

**JOÃO JORGE COMÉDIAS HENRIQUES**

**A AVALIAÇÃO AUTÊNTICA EM EDUCAÇÃO  
FÍSICA – O PROBLEMA DOS JOGOS  
DESPORTIVOS COLECTIVOS**

**Orientadora: Maria do Carmo Clímaco**

**Universidade Lusófona de Humanidades e Tecnologias**

**Instituto de Educação**

**Lisboa**

**2012**

**JOÃO JORGE COMÉDIAS HENRIQUES**

**A AVALIAÇÃO AUTÊNTICA EM EDUCAÇÃO  
FÍSICA – O PROBLEMA DOS JOGOS  
DESPORTIVOS COLECTIVOS**

Tese apresentada para a obtenção do grau de Doutor em Educação no curso de Doutoramento em Educação conferido pela Universidade Lusófona de Humanidades e Tecnologias

Orientadora: Prof<sup>ª</sup> Doutora Maria do Carmo Clímaco

**Universidade Lusófona de Humanidades e Tecnologias**

**Instituto de Educação**

**Lisboa**

**2012**

*Se fosse possível explicar-te tudo não  
precisarias de perceber nada.*

*(Agostinho da Silva)*

À memória de meu pai e de minha mãe  
e à minha filha Inês, permanente na  
minha felicidade e amor incondicional.

Quero, reconhecidamente, agradecer:

A Maria do Carmo Clímaco, a quem reconhecemos a extraordinária sabedoria, que sempre me apoiou com o seu conhecimento e a sua crítica ao longo desta exigente aventura.

Ao Luís Bom, principal companheiro nesta jornada, pela amizade e os momentos de partilha e reflexão.

A Rita Cordovil, pela amizade, conhecimento e companheirismo, por estar sempre presente quando, frequentemente, dela necessitei.

Aos grandes especialistas Luís Almeida, Fátima Fragoso, Carlos Caetano e Joana Maurício, pela amizade e incentivo e, ainda, pela colaboração e capacidade pedagógica, tão necessárias em etapas importantes do trabalho.

Aos professores Jorge Mira, António Lopes, António Palmeira, Sofia Fonseca, António Rosado, Maria Neves, José Alves Diniz, Domingos Fernandes, Joana Nogueira, Paula Castanho, pelo contributo dado em diversas fases do trabalho.

Aos professores e alunos que participaram no trabalho de campo por se terem disponibilizado para integrarem as situações para as quais foram solicitados.

Ao Director do Instituto de Educação, António Teodoro, pelo seu incentivo e exemplo e pela atenção e cuidado na resolução de muitos problemas.

Ao Director da Faculdade de Educação Física e Desporto, Jorge Proença, pela confiança e incentivo no seio da nossa comunidade académica.

A todos os professores que ao longo dos anos têm lutado pela Educação Física das nossas crianças e jovens, em particular, os professores de Educação Física da Escola Secundária Dona Luísa de Gusmão.

A todos os meus colegas e amigos que acreditam no valor do meu trabalho e apoiam o meu trajecto de professor.

À Inês, pelas nossas belas tardes de Sábado e Domingo, olhando o rio e a cidade, e partilhando uns inesquecíveis momentos de pai e filha enquanto estudávamos as nossas matérias.

## Resumo

No esforço de qualificação representado pelas avaliações alternativas, ganha destaque a avaliação autêntica como linha de melhoria do conceito e das práticas de avaliação. Alinhadas com novas perspectivas de ensino dos Jogos Desportivos Colectivos, surgiram formas de avaliação autêntica, alternativas aos testes de habilidades. Contudo, os instrumentos mais utilizados mantêm dificuldades metodológicas, faltando-lhes níveis de jogo que sirvam de referência para a avaliação. Como objectivo geral, este estudo pretende responder ao problema: Como produzir classificações válidas e fiáveis do domínio dos JDC a partir de provas de avaliação autênticas que usam níveis de jogo? O trabalho de campo consistiu em situações de classificação dos níveis de jogo e do desempenho dos jogadores, tendo os grupos experimentais uma instrução para essa classificação distinta do grupo de controlo. Os resultados revelam que o conhecimento dos níveis do programa e a sua utilização fez melhorar, significativamente, as classificações. O grupo que identificou o nível de jogo em primeiro lugar foi o que classificou melhor os jogadores. Nesta abordagem não foram necessários nem recursos especiais, nem muito tempo. Argumentamos em conclusão que, sendo a actividade de grupo a essência da própria matéria, esses resultados decorrem de ser mais correcto partir do grupo para o individual e do jogo para o jogador.

Palavras-chave: avaliação autêntica, educação física, jogos desportivos colectivos, validade, fiabilidade

## **Abstract**

Authentic assessment is an improvement in assessment concept and practice within the shift brought by alternative assessment. In line with new teaching perspectives in Team Sports, some means of authentic assessment have emerged as an alternative to testing skills. However, implementing the most widely used instruments is still a somewhat difficult task, as game levels which may be used as references for assessment are lacking. This study focus mainly the issue: How may valid and reliable ratings in TS be produced from authentic assessment testing using game levels? The field work consisted of game level and players' performance rating situations. Trial groups were specifically trained towards this goal differently from the control group. The results show that the knowledge and identification of programme's performance levels and how they are used in the assessment were instrumental in the significant reduction of assessment errors. The group that first identified the game levels was also the one that best ranked the players. This approach involved neither specific resources nor a lot of time. As group activity is the essence of content in itself, it was deemed more accurate to move from group to the individual and from game to the player.

Keywords: authentic assessment, physical education, team sports, validity, reliability

## **Abreviaturas e Símbolos**

A – nível avançado

AAHPERD – American Alliance for Health, Physical Education, Recreation and Dance

ANOVA – Analysis of Variance

APA – American Psychological Association

CA – cumpre nível avançado

cf - conforme

BC – bolas conquistadas

BJ – bolas jogadas

BP – bolas perdidas

BR – bolas recebidas

CNAPEF – Conselho Nacional das Associações de Professores e Profissionais de Educação Física

E – nível elementar

EF – Educação Física

e. g. – por exemplo

ESDLG – Escola Secundária Dona Luísa de Gusmão

et al – e colaboradores

GEF – Grupo de Educação Física

GPAI – Game Performance Assessment Instrument

I – nível introdução

IE – Índice de Eficiência

i. e. – isto é

IEA – International Association for the Evaluation of Educational Achievement

JDC – jogos desportivos colectivos

KS – Kolmogorov-Smirnov

Ku – Kurtose

NASBE – National Association of State Boards of Education

NRSEF – Normas de Referência para o Sucesso em Educação Física

OCDE – Organisation for Economic Co-operation and Development

PISA – Programme for International Student Assessment

PNEF – Programas Nacionais de Educação Física

SE – Sports Education

Sk – Assimetria

SPSS – Statistical Package for the Social Sciences



SW – Shapiro-Wilk

TG – Tactical Games

TGfU – Teaching Games for Understanding

TSAP – Team Sport Assessment Procedure

VJ – volume de jogo

& - e

## Índice Geral

<b>INTRODUÇÃO .....</b>	<b>14</b>
<b>ENQUADRAMENTO CONCEPTUAL .....</b>	<b>26</b>
<b>1.1. Conceitos e Opções Estruturantes da Avaliação Escolar .....</b>	<b>27</b>
1.1.1. Funções da avaliação .....	29
1.1.2. Qualidades formais da avaliação .....	35
<b>1.2 Problemas tradicionalmente associados à Avaliação Escolar .....</b>	<b>48</b>
<b>1.3. A Avaliação Autêntica no quadro da Avaliação Alternativa .....</b>	<b>61</b>
1.3.1. Características principais na Avaliação Alternativa .....	63
1.3.2. A Avaliação Autêntica .....	70
<b>1.4. Da Avaliação Tradicional à Avaliação Autêntica em Educação Física .....</b>	<b>79</b>
<b>1.5. Discussão da especificidade dos JDC em EF .....</b>	<b>86</b>
<b>1.6. Modelos de ensino baseados numa aprendizagem autêntica dos jogos .....</b>	<b>91</b>
<b>1.7. Avaliação Autêntica nos JDC – “O Estado da Arte” .....</b>	<b>101</b>
1.7.1. O Game Performance Assessment Instrument .....	102
1.7.2. O Team Sport Assessment Procedure .....	107
1.7.3. Outros .....	111
<b>1.8. Como produzir classificações válidas e fiáveis do domínio dos JDC .....</b>	<b>112</b>
1.8.1 Situações de avaliação baseadas no jogo – um contexto autêntico para a avaliação .....	112
1.8.2. Os grandes problemas de avaliação nos jogos .....	114
1.8.3. Como produzir classificações válidas e fiáveis .....	117
1.8.3.1. Do grupo para o indivíduo .....	117
1.8.3.2. Do todo para a parte .....	119
1.8.3.3. Da síntese para a análise, do jogo para o jogador .....	121
<b>CAPÍTULO 2 .....</b>	<b>124</b>
<b>METODOLOGIA E PLANO DE PESQUISA .....</b>	<b>124</b>
<b>2.1. Provas de Avaliação Autêntica no GEF da ESDLG .....</b>	<b>125</b>
<b>2. 2. Procedimentos e critérios adoptados na elaboração dos vídeos de jogo .....</b>	<b>134</b>
2.2.1. Fiabilidade das classificações dos jogadores por um painel .....	135
<b>2.3. O plano de pesquisa .....</b>	<b>137</b>
2.3.1. O estudo-piloto .....	142
<b>2.4. Participantes .....</b>	<b>145</b>
<b>2. 5. Técnica de recolha de dados .....</b>	<b>150</b>
<b>2. 6. Procedimentos .....</b>	<b>151</b>

<b>2. 7. Análise dos dados.....</b>	<b>156</b>
<b>CAPÍTULO 3.....</b>	<b>161</b>
<b>APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS .....</b>	<b>161</b>
<b>3.1. Análise das classificações no conjunto de jogos A (jogos 1, 2 e 3) .....</b>	<b>162</b>
3.1.1. Das diferenças entre as variáveis de caracterização da amostra .....	162
3.1.2. Das diferenças entre grupos.....	163
<b>3.2. Análise das classificações no conjunto de jogos B (jogos 4, 5 e 6) .....</b>	<b>164</b>
3.2.1. Das diferenças entre as variáveis de caracterização da amostra dentro de cada grupo ....	164
3.2.2. Das diferenças entre grupos.....	165
<b>3.3. Análise das diferenças de classificações produzidas por cada grupo nos dois momentos (conjunto de jogos A e conjunto de jogos B).....</b>	<b>177</b>
<b>CONCLUSÕES .....</b>	<b>180</b>
<b>BIBLIOGRAFIA .....</b>	<b>193</b>
<b>APÊNDICES .....</b>	<b>I</b>
<b>ANEXOS .....</b>	<b>X</b>

## Índice de Tabelas

<b>Tabela 1. Estatística descritiva da idade e anos de serviço e análise de frequência do género, estatuto profissional e especialistas em JDC dos grupos de controlo, experimental 1 e experimental 2. ....</b>	<b>149</b>
<b>Tabela 2. Variável grupo no conjunto de jogos a (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>157</b>
<b>Tabela 3. Variável género no conjunto de jogos a (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>157</b>
<b>Tabela 4. Variável tipo de professor no conjunto de jogos a (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>157</b>
<b>Tabela 5. Variável especialista no conjunto de jogos a (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>157</b>
<b>Tabela 6. Variável grupo no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>158</b>
<b>Tabela 7. Variável género no grupo de controlo no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>158</b>
<b>Tabela 8. Variável tipo de professor no grupo de controlo no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>158</b>
<b>Tabela 9. Variável especialista no grupo de controlo no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>158</b>
<b>Tabela 10. Variável género no grupo experimental 1 no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>158</b>

<b>Tabela 11. Variável tipo de professor no grupo experimental 1 no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>159</b>
<b>Tabela 12. Variável especialista no grupo experimental 1 no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>159</b>
<b>Tabela 13. Variável género no grupo experimental 2 no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>159</b>
<b>Tabela 14. Variável tipo de professor no grupo experimental 2 no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>159</b>
<b>Tabela 15. Variável especialista no grupo experimental 2 no conjunto de jogos b (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade. ....</b>	<b>159</b>
<b>Tabela 16. Estatística descritiva (<i>m</i>, <i>dp</i>) da soma de erros por conjunto de jogos (1+2+3) para os grupos de controlo, experimental 1 e experimental 2, e valores da análise de variância univariada relativa à comparação dos erros entre os grupos....</b>	<b>163</b>
<b>Tabela 17. Estatística descritiva (<i>m</i>, <i>dp</i>) da soma de erros por conjunto de jogos (1+2+3 e 4+5+6) para os grupos de controlo, experimental 1 e experimental 2, e valores da análise de variância univariada relativa à comparação dos erros entre os grupos. ....</b>	<b>166</b>
<b>Tabela 18. Estatística descritiva (<i>m</i>, <i>dp</i>) da soma de erros por grupo (controlo, experimental 1 e experimental 2) em cada conjunto de jogos (1+2+3 e 4+5+6), e valores do teste <i>t</i>-student para amostras emparelhadas relativa à comparação dos erros entre os dois momentos.....</b>	<b>177</b>

## Índice de Quadros

<b>Quadro 1. Sistematização dos diferentes tipos de avaliação das aprendizagens que existem no quadro legal português.....</b>	<b>29</b>
<b>Quadro 2. Classificação de jogos. Adaptado de Almond (1986b). ....</b>	<b>87</b>
<b>Quadro 3. Plano de pesquisa. ....</b>	<b>140</b>
<b>Quadro 4. Frequências de remates em salto, golos e progressões no campo no jogo 1. ....</b>	<b>172</b>
<b>Quadro 5. Frequências de remates em salto, golos e progressões no campo no jogo 2. ....</b>	<b>172</b>
<b>Quadro 6. Frequências de remates em salto, golos e progressões no campo no jogo 3. ....</b>	<b>172</b>
<b>Quadro 7. Frequências de remates em salto, golos e progressões no campo no jogo 4. ....</b>	<b>173</b>
<b>Quadro 8. Frequências de remates em salto, golos e progressões no campo no jogo 5. ....</b>	<b>173</b>
<b>Quadro 9. Frequências de remates em salto, golos e progressões no campo no jogo 6. ....</b>	<b>173</b>

## Índice de Figuras

- Figura 1. Média da soma de erros de classificação dos observadores dos três grupos participantes no estudo, no conjunto de jogos a (jogos 1, 2 e 3). As barras de erro representam o desvio-padrão. ....163**
- Figura 2. Média da soma de erros de avaliação dos observadores dos três grupos participantes no estudo, no conjunto de jogos a (jogos 1, 2 e 3) e no conjunto de jogos b (jogos 4, 5 e 6). As barras de erro representam o desvio-padrão. Os asteriscos representam médias significativamente diferentes de acordo com o teste games-howell para  $\alpha = 0.05$ . ....166**
- Figura 3. A soma de erros de avaliação média dos observadores nos três grupos (controlo, experimental 1 e experimental 2) no conjunto de jogos a (jogos 1, 2 e 3) vs. No conjunto de jogos b (jogos 4, 5 e 6). A soma de erros de avaliação média no conjunto de jogos b é significativamente menor que no conjunto de jogos a. Grupo de controlo ( $t(38) = 2.561$ ;  $p = 0.015$ ). Grupo experimental 1 ( $t(57) = 7.824$ ;  $p < 0.001$ ). Grupo experimental 2 ( $t(53) = 12.651$ ;  $p < 0.001$ ). As barras de erro representam o desvio padrão.....178**

## **Introdução**



“Duas equipas e uma bola com todos os seus aspectos características e influências recíprocas constituem um modo excelente de aprendizagem da cooperação, da responsabilidade pessoal, da solidariedade, para além da formação do gosto pela luta e emulação” (Crespo, 1979, p. 64).

Na verdade, os Jogos Desportivos Colectivos [JDC] têm estado muito presentes na formação das crianças e jovens e no quotidiano das escolas, principalmente nas aulas de Educação Física [EF] que é a área disciplinar onde são ensinados. Revendo a bibliografia, Harvey (2006) constata que, em diversos países, uma grande percentagem do tempo do currículo de EF consiste em jogos, nomeadamente 65% no Reino Unido, 50% no Canadá, e 60% na Finlândia em jogos de invasão.

Os estudos de Kenneth Hardman vieram confirmar esta tendência mundial de predominância dos JDC nos quadros de extensão curricular em EF. O autor mostra que 91% dos países inclui os “Jogos de Equipa” no Currículo de EF no Ensino Secundário, enquanto no Ensino Básico, os resultados indicam 96% dos países que o fazem (Hardman, 2008, p.17).

Em termos de tempo de aula atribuído nos Programas aos JDC, os dados dos trabalhos de educação comparada desse autor, mostram que à categoria “Jogos” é dedicado 41% do tempo de aula de EF no Ensino Básico e 43 % no Ensino Secundário (p. 18).

Em Portugal, nos Programas Nacionais de Educação Física [PNEF] (Bom et al., 1989; Jacinto, Carvalho, Comédias & Mira, 2001), podemos identificar com toda a clareza a importância reconhecida ao jogo como expressão da cultura e do tipo de actividade característico da EF. Os objectivos gerais de ciclo assumem a importância do jogo formal como referência para a aprendizagem e a avaliação.

Nos jogos, a forte ligação entre a aprendizagem e a avaliação é afirmada por Harvey (2006), quando considera que ensinar através do jogo e querer que os alunos aprendam como jogar implica avaliar o seu desempenho no jogo.

De facto, o cenário real e autêntico em que se expressam todas as habilidades, conhecimentos e capacidades para jogar é o próprio jogo. Mas o jogo tem características que tornam a tarefa de avaliar muito difícil. Qualquer jogo ocorre num ambiente de grande incerteza e imprevisibilidade (Bom, 1980; Garganta, 1998) onde o desempenho, que por natureza é colectivo, está marcado pela instabilidade das condições em que é avaliado. O jogo é válido, mas os desempenhos estão dependentes do contexto, situação e oponente, e não se repetem, sendo por isso, não fiáveis (Lames & McGarry, 2007).

Todas estas dificuldades existem. Pretendemos aprofundar a dimensão, a natureza e as características das dificuldades, pesquisando uma hipótese de solução. É nesta

clarificação que reside, no essencial, a ideia chave deste estudo, que se orienta no sentido de responder à seguinte pergunta:

Como produzir classificações válidas e fiáveis do domínio dos Jogos Desportivos Colectivos a partir de provas de avaliação autênticas que usam níveis de jogo?

Centrando-nos neste problema, passamos a apresentar genericamente a discussão do Enquadramento Conceptual, desenvolvido no 1º Capítulo, antecedendo a apresentação das opções metodológicas que, a partir da página 24, continuam esta introdução.

Assim, importa desde já reconhecer que se trata de um problema que é transversal a todo o tipo de avaliações, nomeadamente as que já têm existência efectiva no sistema de ensino como as provas sumativas ou as provas de aferição na avaliação interna, ou as que ainda não têm existência legal mas já são defendidas em meios de Educação Física como é o caso das provas de aferição na avaliação externa (Conselho Nacional das Associações de Professores e Profissionais de Educação Física [CNAPEF], 2006; Comédias, 2004, 2005a, 2005b, 2006).

Na perspectiva de que «a avaliação sumativa» e «a avaliação formativa» não são designações para diferentes tipos de avaliação mas antes funções que as próprias desempenham (Harlen & Gardner, 2010; William, 2004) e que as duas lógicas de funcionamento devem operar em conjunto para melhor se complementarem (Bonniol & Vial, 2001; Allal, 2010; Fernandes, 2008), consideramos que a avaliação para fins formativos deverá ser o sustentáculo de qualquer sistema de avaliação.

Partimos, também, do pressuposto que qualquer tipo de avaliação deve respeitar determinadas qualidades formais. Uma boa avaliação deve ser válida e fiável (Nazario, 2005; Mertler & Campbell, 2005), e a compreensão desses conceitos e das relações de dependência entre eles é importante porque, por um lado, são critério para se poder perceber as vantagens e desvantagens de cada avaliação (Harlen, 2007) e, por outro lado, as inferências dignas de confiança dependem deles (Popham, 2003).

Assim, se a informação recolhida respeitar estas qualidades, entendidas como guias de orientação, pode não só ajudar a regular o processo de ensino aprendizagem e permitir tomar decisões relacionadas com o percurso escolar de cada aluno, como, também, caso os dados individuais sejam agregados e processados de forma adequada, servir para apreciar o trabalho de uma escola, departamento, ano de escolaridade, turma, professor, ou, ainda, como é considerado por G. De Landsheere (1997), monitorizar os resultados do sistema educativo, ou de parte dele, de ano para ano.

Só que a avaliação escolar defronta-se com vários problemas que estão identificados e estudados reflectindo convicções, decisões ministeriais e a prática dos professores nas escolas. De facto, predomina uma avaliação que apenas visa classificar os

alunos (Fernandes, 2005; 2007), que está muito pouco relacionada com o processo de ensino aprendizagem e que super valoriza os resultados de final de ano tornando-os, nalguns casos, determinantes para tomar decisões sobre a progressão dos alunos e para apreciar a qualidade do trabalho das escolas (Tucker & Stronge, 2005; Barton, 2008; Popham, 2003).

Diversos autores consideram que, na relação entre instrução e avaliação, há uma tendência pedagógica para se enfatizar fortemente as habilidades discretas e analíticas (Nazario, 2005; Sternberg, 2008; McTighe & O'Connor, 2005). Além disso, entre nós, Fernandes (2005) argumenta que não se verifica nas práticas docentes uma mobilização dos conhecimentos dos alunos para resolver problemas, o que este autor entende ser necessário.

Por outro lado, embora os discursos e teorizações sobre a avaliação coloquem a necessidade de colocar problemas de aplicação de conhecimentos em situações de cooperação entre os alunos, Herrington e Herrington (2006) afirmam que essa não é uma característica que geralmente se observe e que, no ensino, não se encoraja suficientemente a colaboração.

As implicações destas tendências são relevantes ao nível da avaliação pedagógica e tornam-se preocupantes de tal modo que alguns dos principais autores portugueses sobre os temas currículo e avaliação consideram que a maneira como ainda se ensina e avalia é responsável pela falta de competências (Roldão, 2003), e que pode estar relacionada com o abandono precoce e elevadas taxas de retenção (Fernandes, 2007).

Mas, hoje, estamos perante novas exigências curriculares que vão muito para além da memorização. De acordo com Fernandes (2005), preparar para a complexidade e para resolver problemas, utilizar as novas tecnologias, comunicar eficazmente, ter espírito crítico e trabalhar em grupos multiculturais são os grandes desafios e os próprios estudos internacionais sobre avaliação pressionam para que se adoptem padrões curriculares que estejam de acordo com estas novas exigências e que também a avaliação esteja alinhada com essas exigências, ou seja, uma avaliação a preparar os alunos muito mais para o futuro, do que para padrões do passado (Stobart, 2010).

Na tentativa de dar uma resposta mais eficaz a muitos dos problemas associados tradicionalmente à avaliação escolar e em sintonia com as tendências e exigências das sociedades actuais, têm ganho expressão e importância formas de avaliação que se constituem como alternativas ao que tem sido tradicional. Segundo Fernandes (2006), são avaliações mais contextualizadas e integradas no processo de ensino aprendizagem, tendo os alunos um papel mais relevante a desempenhar do que nas provas baseadas nos tradicionais testes padronizados e analíticos. As tarefas correspondem a uma função

formativa da avaliação, estando assim mais orientadas para a melhoria das aprendizagens do que apenas para a classificação.

No esforço de qualificação da avaliação representado pelas avaliações alternativas, ganha destaque uma linha muito promissora de melhoria do conceito e das práticas de avaliação, a chamada avaliação autêntica (Wiggins, 1989; 1990). Na definição de Gulikers, Bastiaens e Kirschner (2004), a avaliação autêntica requer que os alunos usem as mesmas competências ou combinações de conhecimentos, habilidades e atitudes que necessitam aplicar na situação critério na vida real, sendo que o nível de autenticidade é então definido pelo grau de semelhança com a situação critério. Referem que a ênfase deve ser colocada no valor realístico da tarefa e do contexto e sublinham que o resultado da avaliação deve ser um produto ou desempenho de qualidade que possa ser pedido na vida real.

Ao contrário de um exame de Inglês em que a compreensão oral e a expressão oral, tão determinantes na construção da aprendizagem da Língua, não são avaliadas, na avaliação autêntica avaliam-se directamente as competências, respeitando a validade de constructo (Stobart, 2010).

Mas, as provas de avaliação autêntica enfrentam as dificuldades típicas das provas alternativas que são mais difíceis de elaborar e classificar (Nazario, 2005; Fernandes, 2005). Muitas das apreciações que se fazem ao desempenho têm variações muito acentuadas, não havendo garantias que o mesmo desempenho seja avaliado da mesma maneira por avaliadores diferentes (Kellaghan & Madaus, 2003) ou pelo mesmo avaliador em momentos diferentes, como constata Conceição (1994) a propósito de estudos sobre exames, nos itens de resposta aberta. Inclusivamente, as próprias técnicas de validação de provas clássicas que estão baseadas numa abordagem psicométrica já não correspondem à necessidade de se avaliarem competências em produções que são por natureza complexas (De Ketele & Gerard, 2005).

Nas escolas, que desde sempre lidaram com a avaliação (Clímaco, 2005), o inevitável esforço de concertação que é necessário para garantir consistência nas avaliações está comprometido, porque ainda permanece uma ausência de cultura de avaliação nos hábitos, atitudes e comportamentos. A construção do acordo entre quem avalia é uma questão social na escola e implica mudanças, pois como é considerado por Clímaco (2005), Fernandes (2005) e Fernandes, Neves, Campos e Lalande (1996) a avaliação, na maioria das situações, é uma actividade muito solitária na forma como é vivida, pensada e praticada.

Mesmo quando o avaliador se encontra sozinho precisa de uma estrutura conceptual construída em grupo, ou seja, a intersubjectividade tem de estar resolvida. Para o treino de aferição de critérios e de avaliação é necessário tempo de trabalho colectivo

entre os professores. A avaliação não se muda por decreto que é o que se tem feito e que tem como resultado que tudo fique na mesma (Pacheco, 1994). É na mudança da «cultura de escola» que se deve apostar, já que, de acordo com Barroso (2005), é ela que pode relativizar a chamada «cultura escolar», que corresponde aos “valores, hábitos e comportamentos, transmitidos pela forma escolar de educação a partir de determinações exteriores” (p. 56). Para que a mudança aconteça os professores precisam de tempo para trabalhar em conjunto (Darling-Hammond, 2004) e de uma nova cultura de trabalho em equipa que garanta a validade e fiabilidade das avaliações, permitindo que se façam inferências válidas dos seus resultados.

Na EF, também as abordagens de avaliação são o reflexo da forma como os currículos têm sido construídos e da influência de correntes pedagógicas tradicionais (Hay, 2006). Essas abordagens são marcadas pela falta de autenticidade e de contextualização das técnicas tradicionais de avaliação, que tratando de forma separada os domínios cognitivo, afectivo e psico-motor, tendem a avaliar de forma muito estrita as suas características (Wood, 2003).

No caso específico dos JDC, os instrumentos mais utilizados para avaliar a capacidade de jogo são os testes de habilidades, inspirados em testes padronizados como os da American Alliance for Health, Physical Education, Recreation and Dance [AAHPERD] (Kirkendall, Gruber & Johnson, 1987),

De acordo com Gréhaigne, Richard e Griffin (2005), na abordagem tradicional o objectivo é dominar um conjunto de habilidades técnicas para mais tarde serem aplicadas em situação de jogo e só depois introduzir os aspectos de natureza táctica. Mas, como sublinham Kirk e McPhail (2002), raramente o professor estabelece ligações entre a técnica e o contexto de jogo.

“Though rules and strategies for playing games have changed over the years, methods used to teach and assess games have changed very little” (Oslin, Mitchell & Griffin, 1998, p. 231).

De facto, muito do que se sabe sobre avaliação nos JDC apresenta-se sobre a forma de testes de habilidades, que tendem a avaliar apenas uma componente do jogo, a execução da habilidade discreta. São, portanto, testes que estão focados nos aspectos técnicos (Gréhaigne et al., 2005). Parte-se do princípio que as pontuações dos testes de habilidade representam a habilidade para utilizar as habilidades no jogo, quando de facto, não representam o seu uso adequado, ou seja, quando e onde usá-las (Oslin et al., 1998).

Mas, também, na EF e em particular nos JDC, a avaliação autêntica ganha expressão. De acordo com Gréhaigne et al. (2005), os investigadores em pedagogia do desporto concordam com a importância da avaliação autêntica no processo de ensino-

aprendizagem nos JDC. Harvey (2006) preconiza, mesmo, mais avaliação autêntica do jogo em EF. Mais especificamente, nesta área disciplinar, estas avaliações são baseadas no desempenho, são administradas em contextos de vida real semelhantes aquele em que as competências se devem expressar (Wood, 2003; Siedentop & Tannehill, 2000) e aproveitam as situações realistas (Harvey, 2006).

“A avaliação autêntica reporta-se a desempenhos contextualizados, procurando-se um alinhamento da instrução com a avaliação. Deste modo, as práticas de avaliação não suspendem o processo de aprendizagem, são, antes de mais, oportunidades para os alunos aprenderem e exercitarem os resultados desejados e receberem feedback sobre o desenvolvimento da sua aprendizagem” (Graça & Mesquita, 2007, p. 410).

Na perspectiva da avaliação autêntica, interessa-nos analisar e estudar especificamente os jogos de equipa, aqueles que são jogados por duas equipas. As equipas opõem-se de forma a marcar pontos ou impedir que o opositor marque. Para que o objectivo se concretize, os jogadores relacionam-se entre si, movimentando a bola ou um projectil similar, de acordo com um conjunto de regras (Lames & McGarry, 2007; Garganta, 2009). Neste tipo de jogos, jogar bem significa escolher a acção certa no momento certo, desempenhá-la eficientemente e fazê-lo com frequência durante o jogo (Gréhaigne et al., 2005; Gréhaigne, Godbout & Bouthier, 2001).

As características principais dos jogos de equipa e o próprio conceito sobre o que é um bom desempenho nos jogos desportivos colectivos, tornam, assim, a disputa ou a separação entre a técnica e a tática, um vazio sem sentido. As técnicas fazem parte do reportório das possíveis escolhas do jogador, “são simultaneamente acções no sentido elementar do termo, e elementos de acções complexas” (Mahlo, 1969/n. d., p.80). São acções porque os processos de natureza técnica são transformados pelo pensamento, quando ele lhes dá um conteúdo tático (Téodorescu, 1977).

Nos últimos trinta anos verificou-se uma viragem acentuada na perspectiva do ensino dos jogos, que se traduziu numa tentativa para permitir uma aprendizagem mais aproximada à actividade cultural referente e, por isso, mais autêntica. Em Portugal, Barreto et al. (1980) preconizaram que o método de ensino se centre no jogo e no aluno. O jogo passou a ser visto como a referência central para o processo de aprendizagem. Da ênfase no ensino das técnicas isoladas que depois eram aplicadas no jogo, passou-se para a capacidade de jogar através da compreensão tática do jogo, resolvendo os problemas que ele coloca (Graça & Mesquita, 2007).

Em todo o mundo, a necessidade tão reclamada de maior autenticidade, particularmente na sub-área dos JDC trouxe, novas formas de ensinar como o Teaching Games for Understanding [TGfU] (Bunker & Thorpe, 1982), o Tactical Games [TG] (Griffin, Mitchell & Oslin, 1997; Mitchell, Oslin & Griffin, 2003), o Sports Education [SE] (Siedentop,

1994), a Aprendizagem Situada (Kirk & McPhail, 2002), a Abordagem Ecológica (Chow, Davids, Button, Shuttleworth, Renshaw & Araújo, 2007), o Modelo de Competências nos Jogos de Invasão (Graça et al., 2003).

Por outro lado, a necessidade de alinhar uma avaliação autêntica com estas novas formas de ensinar trouxe, como consequência, novas formas de avaliar alternativas aos tradicionais testes de habilidades sem contextualização. De entre todos os esforços que foram desenvolvidos para procurar avaliar situações de jogo autênticas, destacam-se, na literatura publicada, os dois mais importantes instrumentos de avaliação o Game Performance Assessment Instrument [GPAI] (Griffin et al., 1997; Oslin et al., 1998) e o Team Sport Assessment Procedure [TSAP] (Gréhaigne & Godbout, 1998; Gréhaigne, Godbout & Bouthier, 1997).

No entanto, as conclusões de vários estudos apontam para uma dificuldade na utilização destes instrumentos. Faltam níveis de jogo, em que o desempenho esteja categorizado, que possam servir de referência para a avaliação (Harvey, 2006; Tallir, Musch, Lannoo & Van de Voorde, 2004). Existe muito pouco, na matéria publicada, sobre padrões ou descritores que faça a separação entre níveis (Hay, 2006).

Memmert e Harvey (2008), tentando melhorar o GPAI, enfrentaram a dificuldade de afirmar quando uma acção é ou não adequada, especialmente quando se codificam comportamentos sem bola. As diferentes visões, que vários codificadores podem ter quando categorizam os acontecimentos, são fonte de divergências e podem pôr em causa a utilidade do instrumento.

Mas a avaliação também deve ser eficiente nos custos e no tempo dispendido (Allen, 2002). Estamos perante um problema de economia de esforços e de procedimentos para se garantir validade e fiabilidade na avaliação. Nos JDC, os instrumentos de avaliação que permitem avaliar situações de jogo autênticas têm-se revelado muito trabalhosos, consumindo muito tempo. Clama-se por instrumentos que permitam aos professores avaliar enquanto ensinam e observam os seus alunos em actividade (Harvey, 2006).

Assim, enquadrados por esta perspectiva conceptual, importa então explicitar quais são as grandes questões da investigação.

Se a situação de jogo é o referente da aprendizagem a avaliar, como utilizar, então, o jogo na sua forma característica, 5X5 e 7x7 no Andebol, por exemplo, como referente e situação de prova que permita a classificação individual do desempenho do aluno?

Essa possibilidade tem sido discutida por vários autores e tem sido objecto de diversas propostas, sendo o GPAI e o TSAP as mais estudadas. Mas continuam a colocar-se dificuldades a partir da aplicação e pesquisas sobre estas propostas, as quais se concentram na necessidade de sistematização do próprio jogo em termos de níveis de

interpretação prática. Tais níveis são uma preocupação metodológica e didáctica, relevante para a avaliação, a partir principalmente de 1980 (Barreto et al., 1980), a qual veio a reflectir-se no modelo de elaboração de Programas Nacionais de Educação Física (Bom et al., 1989).

Neste modelo curricular, formulado e aplicado nos PNEF, todas as matérias estão especificadas por níveis de prática, em progressão – Introdução, Elementar e Avançado<sup>1</sup>, com destaque para os JDC.

Nestas condições, dispondo da sistematização em que se definem níveis de jogo, essa definição pode traduzir-se em classificações fiáveis do desempenho individual, observado na forma característica da prova de avaliação?

Isso depende, obviamente, da capacidade do observador em atribuir significado aos comportamentos evidenciados, nos termos próprios da especificidade da definição de cada nível. Admite-se que a distinção de conteúdo dos níveis constitui uma sistematização válida da progressão da competência do jogador, na sua aprendizagem, pela forma como foram elaboradas e revistas as especificações das matérias dos PNEF (Bom et al., 1989, 1990; Jacinto et al., 2001) através da colaboração de painéis de especialistas e da consulta e acompanhamento dos professores envolvidos na aplicação dos Programas, em diversas escolas. Esta definição de níveis é um dos traços mais importantes deste modelo curricular, como foi caracterizado por Carreiro da Costa e Piéron (1997, p. 234).

Mas a utilização dos níveis para distinguir e avaliar etapas distintas de aprendizagem, depende também do observador, em termos do conhecimento do jogo, do modelo e das etapas de aprendizagem, a didáctica específica dos JDC, e da sua capacidade de observação dos comportamentos e as qualidades da dinâmica global de jogo, isto é a fluidez e o balanço ofensivo / defensivo entre as equipas que permita distinguir o nível evidenciado.

Estas considerações básicas colocam a necessidade de formação específica dos avaliadores no conhecimento e da análise da matéria, o jogo de Andebol, ou outra matéria dos PNEF, mas não só: formação específica e treino, também, de procedimentos de observação e registo que permitam formalizar as situações de prova e a notação do que se observa, em termos de classificações fiáveis dos desempenhos dos jogadores, que sejam representativas dos respectivos níveis de domínio da matéria.

---

<sup>1</sup> Introdução, onde se incluem as habilidades, técnicas e conhecimentos que representam a aptidão específica ou preparação de base («fundamentos»);  
Elementar, nível em que se discriminam os conteúdos que representam o domínio (mestria) da matéria nos seus elementos principais e já com carácter mais formal, relativamente aos modelos de prática e organização da actividade cultural referente;  
Avançado, que estabelece os conteúdos e formas de participação nas situações típicas da actividade referente, correspondentes ao nível superior que poderá ser atingido no âmbito da disciplina de Educação Física.



Tal formação específica para a avaliação das matérias é um requisito fundamental para que as notações dos professores possam traduzir-se em avaliações, sejam de carácter formativo, sejam na função sumativa.

Neste quadro, interessa-nos estudar a organização de procedimentos de avaliação da situação formal de jogo e também o treino dos observadores, procurando verificar se a observação em tais condições, com referência aos níveis de jogo definidos nos PNEF, permite obter notações, ou classificações objectivas, ou seja, em que as notas atribuídas por diferentes avaliadores sejam coerentes entre si e efectivamente representativas das qualidades da dinâmica do jogo e dos desempenhos observados.

Assim, o nosso problema e as nossas questões de investigação concentram-se nas seguintes hipóteses operacionais de trabalho:

- Não há diferença entre as classificações do desempenho dos jogadores em prova autêntica de andebol, produzidas, por um lado, pelos professores com instrução prévia sobre os níveis de andebol dos PNEF e, por outro lado, as classificações produzidas por professores sem essa instrução.
- Não há diferença entre as classificações do desempenho dos jogadores em prova autêntica de andebol, produzidas, por um lado, pelos professores com instrução prévia para identificar o nível de jogo antes de classificar os jogadores e, por outro lado, as classificações produzidas por professores sem essa instrução.

Para desenvolver o estudo focado nestas hipóteses, parece-nos que há três condições metodológicas fundamentais que é necessário garantir, que estão desenvolvidas no 2º Capítulo, Metodologia e Plano de Pesquisa, e que formulamos do seguinte modo:

Primeiro, só é possível tentarmos verificar a validade e fiabilidade de procedimentos de avaliação autêntica nos JDC se dispusermos de uma referência empírica dessa abordagem de avaliação pedagógica, em que nos possamos basear para definir um protocolo experimental ou *quasi* experimental, a aplicar a uma amostra mais alargada. Necessitamos, pois, de um modelo ou padrão de avaliação de uma escola, suficientemente experimentado e apurado, permitindo operacionalizar as regras segundo as quais se irá submeter uma amostra alargada de professores, às tarefas de avaliação autêntica dos JDC. Tais tarefas terão de ser baseadas na discriminação de níveis de jogo, em condições controladas de (i) formação no domínio das matérias e respectivos níveis programáticos e ainda (ii) de treino de observação das qualidades do jogo e do desempenho dos jogadores, segundo esses níveis.

Esse caso padrão, de uma escola em que se pratique a avaliação autêntica em EF, segundo os níveis e objectivos dos PNEF, é o protocolo de avaliação desenvolvido pelos professores do Grupo de Educação Física [GEF] da Escola Secundária de D. Luísa de

Gusmão [ESDLG] (GEF-ESDLG, 2010). Este protocolo é descrito e analisado no capítulo de Metodologia e Plano de Pesquisa, onde argumentamos as condições e requisitos que nos permitem organizar o desenvolvimento da nossa pesquisa com base nas regras e procedimentos de avaliação dos JDC desta Escola.

Em segundo lugar, precisamos de objectos de avaliação que possam ser classificados em termos de definição de uma «regra de ouro» (Kupermintz, 2004), ou seja um padrão de classificação válida e fiável, antes de serem submetidos ao escrutínio dos sujeitos da amostra, de modo que tenhamos a referência de notação da situação de jogo e de cada jogador, em relação à qual sejam apreciadas as classificações produzidas por estes sujeitos.

Tais objectos de avaliação, que possam ser reproduzidos e avaliados repetidamente, nas mesmas circunstâncias, são muito comuns na generalidade das disciplinas do currículo, com desempenhos que se expressam em redacções, ou que são representados em suporte físico duradouro e reproduzível como papel, quadros, figuras, maquetas e outros. Em EF os desempenhos são performances comportamentais, transitórias e circunstanciais, que pela sua natureza irrepetível, requerem o uso de filme ou gravação em vídeo como única solução para se obter um registo permanente do desempenho e das situações de prova.

Esta característica específica da EF foi discutida por Siedentop (1991) no que designou por «registos permanentes». Assim, para o nosso estudo a solução para estes produtos são os filmes elaborados a partir de provas de JDC utilizados pelos professores da ESDLG para aferir a aplicação de critérios entre subgrupos de docentes de EF, procedimento corrente destes professores para verificarem, em cada ano, a fiabilidade das suas classificações, desde o ano lectivo de 2002/2003.

Em terceiro lugar, desenhámos e seguimos um plano de pesquisa *quasi* experimental segundo o modelo designado como «grupo de controlo não equivalente» (Tuckman, 2000; Thomas, Nelson & Silverman, 2007).

Assim, adaptámos o protocolo de avaliação dos JDC dos professores do GEF da ESDLG, especificamente para a aferição de critérios de classificação, e essa adaptação foi o tratamento base para distinguir o grupo de controlo e os grupos experimentais na nossa amostra.

O trabalho de campo consistiu portanto em situações de classificação dos níveis de jogo e do desempenho dos jogadores, sendo os grupos experimentais seguidos no processo de formação e treino para essa classificação, enquanto o grupo de controlo realizava a notação dos jogo e dos jogadores após um tratamento tipo «placebo» (Thomas et al., 2007),

ou seja, recebia informação irrelevante em termos de influência na qualidade das classificações em causa.

No segundo capítulo, para além do desenvolvimento das condições metodológicas fundamentais anteriormente consideradas, apresentam-se também outras opções metodológicas, englobando a descrição do estudo-piloto e suas principais conclusões, a caracterização da amostra, os procedimentos de recolha de dados utilizados durante pesquisa e a descrição e justificação do processo de intervenção aplicado aos diferentes grupos de observadores/avaliadores. O capítulo termina com a apresentação dos procedimentos de análise dos dados.

No terceiro capítulo apresentam-se e discutem-se os resultados. Expõe-se o comportamento da amostra nas variáveis estudadas. Tomando como referência as duas hipóteses de estudo, os resultados são apresentados e discutidos em função das diferenças entre as variáveis de caracterização da amostra e das diferenças entre os grupos (controlo, experimental 1 e experimental 2), nos dois momentos, antes e depois da intervenção.

Finalmente apresentam-se as conclusões do estudo, destacando os elementos essenciais retidos da discussão dos resultados. Nestas conclusões são, também, referidas as implicações para a avaliação dos alunos e as implicações na organização das escolas que este estudo sugere e recomenda. São, ainda, apresentadas pistas de estudo para futuras investigações.

Importa ainda notar que para as citações e referências bibliográficas que constam do texto serão utilizadas as normas da APA.

# **Capítulo 1**

## **Enquadramento Conceptual**

### **1.1. Conceitos e Opções Estruturantes da Avaliação Escolar**

A avaliação das aprendizagens que, de acordo com Fernandes (2005), é “todo e qualquer processo deliberado e sistemático de recolha de informação, mais ou menos participado e interactivo, mais ou menos negociado, mais ou menos contextualizado, acerca do que os alunos sabem e são capazes de fazer numa diversidade de situações” (p. 16), tem impacto nos alunos, nos professores, no currículo, nos decisores políticos e na opinião pública. A natureza desse impacto depende muito dos objectivos subjacentes a cada tipo de avaliação, já que a avaliação das aprendizagens pode ser utilizada com diferentes propósitos.

Há avaliações que recolhem e interpretam dados sobre o desempenho individual e que servem para tomar decisões que dizem, exclusivamente, respeito a cada aluno e há avaliações que agregam os dados do desempenho individual em grupos para depois de devidamente processados permitirem avaliar programas, organizações, anos, turmas, escolas, sistemas, etc.

Harlen (2007) faz a distinção entre a avaliação «assessment», quando a recolha de informação, as apreciações e as decisões dizem respeito ao desempenho dos alunos na relação com determinadas metas de aprendizagem e a avaliação «evaluation», quando a recolha de informação, as apreciações e as decisões estão relacionadas com sistemas, programas, materiais, procedimentos e processos. Esta autora considera que os dados individuais dos alunos, se forem agregados em grupo, podem ser utilizados neste segundo tipo de avaliação, e refere que os dois processos de avaliação são similares, embora difiram quanto ao que está na base dos juízos.

Quer a utilização de dados individuais de cada aluno, quer a utilização de dados de grupos de alunos, podem ser encontradas tanto na avaliação interna, como na avaliação externa.

Assim, a partir da distinção feita por Harlen, e tomando em consideração o quadro legal português (Decreto-Lei nº 6/2001 de 18 de Janeiro; Despacho Normativo nº 1/2005 de 5 de Janeiro) interessa-nos sistematizar os diferentes tipos de avaliação que dizem respeito à avaliação das aprendizagens, já que o problema que pretendemos resolver com o contributo deste estudo é transversal a todos eles, a saber:

1 - A Avaliação Interna que está sob o controlo de cada escola, dentro dos limites da sua própria política de avaliação. Na avaliação interna considera-se a avaliação que tem objectivos relacionados com o desempenho de cada aluno e a avaliação que tem objectivos relacionados com agregar dados dos desempenhos de grupos de alunos.

Na avaliação que tem objectivos relacionados com o desempenho de cada aluno incluem-se:

a) A avaliação para fins formativos, que ajuda os alunos a aprender (Harlen, 2007), é “interactiva, centrada nos processos cognitivos dos alunos e associada aos processos de feedback, de regulação, de auto-avaliação e de auto-regulação das aprendizagens (Fernandes, 2006, p. 23);

b) A avaliação com fins sumativos realizada ao nível de turma, que é controlada pelo professor titular da turma recolhendo informação em provas, testes, tarefas especiais ou no trabalho do dia-a-dia e que é enquadrada pelos critérios de avaliação aprovados em Conselho Pedagógico que são os mesmos para todos os alunos de um ano/disciplina.

Na avaliação que tem objectivos relacionados com agregar dados dos desempenhos de grupos de alunos consideram-se:

c) As provas de aferição realizadas ao nível do Grupo/Departamento que permite em cada escola os órgãos de gestão pedagógica monitorizarem áreas particulares, com critérios de avaliação que são os mesmos para todos os alunos de um ano/disciplina. Este tipo de provas existe, actualmente, em escolas que decidiram manter as provas globais como instrumento de aferição de conhecimentos (Despacho Normativo nº 5/2007 de 10 de Janeiro).

2 - Avaliação Externa que está sob o controlo dos Serviços Centrais do Ministério da Educação ou é administrada por si, dentro dos limites da sua política de avaliação ou das entidades que promovem os estudos internacionais sobre avaliação. Na avaliação externa, também se considera a avaliação que tem objectivos relacionados com o desempenho de cada aluno e a avaliação que tem objectivos relacionados com agregar dados dos desempenhos de grupos de alunos,

Na avaliação que tem objectivos relacionados com o desempenho de cada aluno incluem-se:

a) Os exames nacionais, que são de natureza sumativa e, certificam e seleccionam.

Na avaliação que tem objectivos relacionados com agregar dados dos desempenhos de grupos de alunos consideram-se:

b) As provas de aferição, que monitorizam algumas disciplinas em todo o sistema permitindo compararem resultados de ano a ano;

c) As provas dos estudos internacionais de avaliação, do Programme for International Student Assessment [PISA] e da International Association for the Evaluation of Educational Achievement [IEA], que comparam os desempenhos de alunos da mesma idade nos países participantes.

Nas provas de aferição realizadas ao nível do grupo/departamento de escola – provas globais, nas provas de aferição realizadas ao nível de todo o sistema educativo e nas provas dos estudos internacionais de avaliação a informação recolhida na avaliação não é

usada para tomar decisões que directamente tenham repercussões em cada aluno. No entanto, na perspectiva de Harlen (2007), a repercussão deste tipo de provas pode sentir-se através do impacto que elas têm no ensino e no currículo.

Quadro 1. Sistematização dos diferentes tipos de avaliação das aprendizagens que existem no quadro legal português.

	<b>Objectivos relacionados com o desempenho de cada aluno</b>	<b>Objectivos relacionados com agregar dados dos desempenhos de grupos de alunos</b>
<b>Avaliação Interna</b>	- Avaliação para fins formativos - Avaliação com fins sumativos realizada ao nível da turma	- Provas de Aferição realizadas ao nível do grupo / departamento
<b>Avaliação Externa</b>	- Exames Nacionais	- Provas de Aferição - Provas dos estudos internacionais de avaliação (PISA; IEA)

Sistematizados os diferentes tipos de avaliação importa, em primeiro lugar, esclarecer as diferentes funções da avaliação, discutindo a forma como coerentemente se deverão relacionar, em segundo lugar, discutir as qualidades formais que qualquer avaliação deve respeitar e, em terceiro lugar, discutir a forma como essas qualidades devem ser utilizadas nos diferentes tipos de avaliação sistematizados anteriormente.

#### **1.1.1. Funções da avaliação**

Neste ponto vamos esclarecer as funções da avaliação que estão consignadas no quadro legal português (Decreto-Lei nº 6/2001 de 18 de Janeiro; Despacho Normativo nº 1/2005 de 5 de Janeiro), concretamente a função diagnóstica, a função formativa e a função sumativa da avaliação.

Partimos do pressuposto que a avaliação formativa e a avaliação sumativa não são diferentes tipos de avaliação mas, que servem diferentes fins, sendo mais adequado substituir os termos avaliação formativa e avaliação sumativa por avaliação para fins formativos e avaliação com fins sumativos (Harlen & Gardner, 2010). Assim, como estas designações se aplicam não às próprias avaliações mas às funções que servem, a mesma avaliação pode ser formativa e sumativa (William, 2004).

De facto, tal como em outras áreas do currículo onde os processos de avaliação podem ser utilizados para diferentes fins, também na EF a informação proveniente de uma mesma situação de avaliação nos JDC tem o potencial para ser utilizada directamente no apoio à melhoria das aprendizagens e da qualidade do jogo ou, então, para certificar as aprendizagens realizadas num dado período de tempo.

Vamos, também, discutir o conflito sem sentido que parece opor a avaliação formativa à avaliação sumativa e que tem condicionado, segundo Bonniol e Vial (2001) a

possibilidade destas duas lógicas de funcionamento se poderem misturar com inteligência para operarem juntas e se complementarem.

#### *A avaliação formativa*

O termo formativo é usado para identificar a avaliação que promove aprendizagem, sendo a informação recolhida utilizada para planear os passos seguintes da aprendizagem e, ao mesmo tempo, saber como fazê-lo (Harlen, 2007). Aliás, avaliação que não ajude a aprendizagem não pode ser chamada de formativa (Harlen & Gardner, 2010).

Alguns autores designam-na como «avaliação para as aprendizagens», sendo os seus objectos, justamente, os processos de aprendizagem (Fernandes, 2008).

Esta avaliação é contínua, ocorre no decurso da instrução, fornece feedback específico a professores e alunos com o objectivo de orientar o ensino para melhorar a aprendizagem, e inclui métodos formais e informais (McTighe & O'Connor, 2005). Pacheco (1994) reconhece, mesmo, ser o feedback fornecido pela avaliação formativa que permite sua própria regulação e considera que são as interpretações que se fazem que podem conduzir a alterações didácticas ou curriculares. Para Clímaco (2005), a avaliação é, precisamente, formativa quando produz informação que permita introduzir correcções na orientação e execução de um programa.

Na tradição francófona, utiliza-se o termo «formadora» para distinguir da avaliação formativa quando, segundo Bonniol (2001b), a regulação se refere prioritariamente às estratégias do professor. Para Vial (2001b), a avaliação formativa não deve ser o que o professor faz para que os alunos aprendam, mas o que é que fazem os professores e os alunos em relação a isso, e como os professores podem ajudar os alunos a aprender. De acordo com Fernandes (2005, 2008)<sup>2</sup>, a tradição francófona vê a avaliação formativa como fonte de regulação dos processos de aprendizagem, tendo os alunos um papel destacado e autónomo. Já para os anglo-saxónicos, o papel primordial na avaliação formativa é o do feed-back que se confunde com ela e é a essência da orientação e do apoio aos alunos.

Por outro lado, Pacheco (1994), apoiando-se noutros autores, compara o behaviorismo e o cognitivismo, referindo que na estratégia behaviorista há uma verificação mecanicista do aproveitamento dos alunos e actividades de remediação. Calcula-se a distância a que se está dos objectivos, transformando o formativo numa prática meramente classificativa de recolha e interpretação dos dados, mantendo os alunos numa atitude passiva. Para o autor, na estratégia cognitivista, o diagnóstico e a orientação são

---

<sup>2</sup> Para este autor, na tradição francófona o auto-controlo, a auto-avaliação, a auto-regulação são os processos cognitivos e metacognitivos destacados, estando a auto-avaliação quase confundida com a avaliação formativa e o papel do professor reduzido ao mínimo. Na tradição anglo-saxónica está tudo mais dependente do professor que planifica, selecciona tarefas e estratégias de avaliação, havendo uma relação estreita entre a avaliação formativa e o currículo.



permanentes, adaptando-se o ensino ao ritmo de cada um, havendo lugar ao imprevisto e ao inesperado. Procura saber-se o que está na base das dificuldades, tomando o aluno consciência dessas dificuldades e não apenas indicar actividades de remediação.

Concluimos, de acordo com Fernandes (2006), que a avaliação também só é formativa quando os alunos tomam consciência das diferenças entre o que são e o que podem vir a ser e estão dispostos a percorrer o caminho que elimina essas diferenças. Como refere Crahay (2002), só há educação quando um indivíduo quer tornar-se noutro que não é, ou seja, passar do estado actual a um outro melhor.

#### *A avaliação diagnóstica*

Harlen (2007) considera que a avaliação formativa inclui a avaliação diagnóstica. Segundo McTighe e O'Connor (2005), esta avaliação é algumas vezes conhecida como pré-avaliação, precedendo a instrução e fornecendo informação para apoiar o planeamento e orientar a diferenciação da instrução. É um processo usado para descobrir qual o conhecimento anterior e nível de habilidade dos alunos, identificando as suas dificuldades, os interesses e preferências.

Na EF, a função diagnóstica da avaliação é alargada, nos seus horizontes, para uma visão que passou a ser prognóstica e associada à avaliação inicial. Luís Bom e a sua equipa reconhecem isso mesmo ao referirem nas orientações metodológicas dos PNEF (1989) e confirmadas na Revisão Curricular de 2001, que a primeira etapa do ano lectivo deve ser uma etapa de Recepção e Orientação dos alunos para o Sucesso do qual faz parte integrante um período de Avaliação Inicial. Para Carvalho (1994), interpretando as orientações dos PNEF (Bom et al., 1989) este período de avaliação inicial deve ter as seguintes características:

- Ser um período de tempo alargado, de quatro ou cinco semanas, em que se recolhem dados sobre a forma como os alunos aprendem de maneira a prognosticar o seu percurso de desenvolvimento;
- Ser um período em que se abordam as matérias seleccionadas para esse ano, respeitando as orientações do programa nacional;
- Ser um período em que se viva um clima de constante desafio e não apenas «exame», com contexto de aulas, onde se dá essa avaliação, semelhante ao das aulas «normais».

De facto, concordamos que para planear uma viagem ao longo do ano agradável e útil, dos alunos com o seu professor, em direcção a objectivos de aprendizagem orientados para competências efectivas, é necessária uma avaliação inicial prolongada no tempo. Esta avaliação, mais do que diagnosticar os conhecimentos, habilidades e atitudes no início do

ano lectivo, deve permitir prognosticar as metas de aprendizagem a alcançar e o percurso de desenvolvimento para lá chegar.

#### *Avaliação sumativa*

A avaliação sumativa mede e classifica o grau de consecução no final de um processo e certifica o resultado (Pacheco, 1994). Refere-se aos efeitos de um programa (Clímaco, 2005). Por contabilizar o que os alunos aprenderam no final de um período de instrução, é principalmente traduzida em classificações ou pontuações (McTighe & O'Connor, 2005).

Alguns autores designam-na como «avaliação das aprendizagens», sendo os seus objectos, precisamente, os resultados das aprendizagens dos alunos (Fernandes, 2008). O autor considera que, embora possa ter uma utilização formativa, normalmente, fica-se pela atribuição de classificações que são registadas e tornadas públicas, representando a síntese das aprendizagens dos alunos.

De acordo com Harlen e Gardner (2010), estudos sobre as práticas de avaliação revelam que a avaliação sumativa tem grande impacto na aprendizagem, embora seja a forma como se utiliza a informação recolhida que determina a influência positiva ou negativa deste impacto.

Por outro lado, como traduz a soma das conquistas realizadas pelos alunos num dado período de tempo, fornece informação para todos os que têm interesse nelas, principalmente os próprios alunos, as famílias, outros professores, as instituições do ensino superior e os empregadores (Harlen, 2007). Para McTighe e O'Connor (2005) é, precisamente, por contar para progredir ou certificar que comanda a atenção de alunos e pais.

#### *Avaliação formativa versus avaliação sumativa – um conflito sem sentido*

Anteriormente quando se falava de avaliação, isso, era sinónimo de avaliação sumativa, de notas e testes, de controlo e balanço. Depois a avaliação formativa passou a ser um valor absoluto, desacreditou-se a nota e passou a ser muito mal visto tudo o que se relacione com normas e balanços (Bonniol & Vial, 2001).

Para Bonniol (2001a), esta oposição entre avaliação formativa e sumativa é falaciosa, não havendo nenhuma oposição mas uma continuidade porque a avaliação formativa também está interessada em resultados, já que é a partir de resultados parciais que se podem avaliar procedimentos e processos. Para este autor há apreciações que somam mais que as notas e a avaliação formativa pode, mesmo, levar a uma nota.

Segundo Pacheco (1994), o conflito entre as diferentes noções de avaliação acontece porque há um alinhamento, ou com o paradigma quantitativo ou com o qualitativo. No entanto, para Bonniol (2001a), é absurdo pensar que a avaliação sumativa só utiliza

critérios quantitativos e objectivos e a avaliação formativa critérios qualitativos e subjectivos e, sublinha que, os critérios quantitativos só assumem sentido em função dos critérios qualitativos dos quais são desdobramentos. Para este autor, para quantificar um critério, terão de seguir-se muitos passos de codificação ou tradução da subjectividade em indicadores quantitativos, distinguindo-se as duas lógicas da avaliação porque a função é diferente.

A avaliação deve articular o que está disponível, mesmo sendo contrário ou contraditório, não se reduzindo a um modelo ou solução, mas multiplicando inteligentemente as várias possibilidades (Bonniol & Vial, 2001).

De acordo com Allal (2010), embora a avaliação para aprender esteja principalmente preocupada com a função formativa da avaliação, pode, contudo, incorporar formas de avaliação sumativa que são concebidas para exercer uma influência positiva na forma como os alunos abordam a aprendizagem. Para esta autora, a continuidade entre a avaliação formativa e a avaliação sumativa desenvolve-se se as duas estiverem alinhadas com os objectivos do currículo subjacentes ao ensino e à aprendizagem na sala de aula e se a avaliação sumativa fornecer aos alunos um feedback de alta qualidade sobre os resultados da aprendizagem.

Para Fernandes (2005) uma abordagem da avaliação das aprendizagens para classificar ou certificar é diferente de uma abordagem da avaliação para que os alunos aprendam, sendo que cada uma tem procedimentos que lhe são específicos. No entanto, este autor argumenta que deve coordenar-se e articular-se as modalidades de avaliação, referindo que as avaliações certificativas da responsabilidade dos professores podem e devem fazer uso das informações recolhidas durante o processo de avaliação formativa, sublinhando que há sistemas educativos que procuram articular as funções da avaliação formativa com as funções da avaliação certificativa, compatibilizando-as. Também Clímaco (2005) considera que as duas modalidades estão muitas vezes associadas e nem sempre se distinguem nitidamente, referindo ser frequente a avaliação sumativa desempenhar uma função proactiva ao serviço da tomada de decisões.

Harlen (2006) considera que as duas modalidades de avaliação são dimensões do mesmo constructo. Estabelecem um conjunto de possibilidades de avaliação com diferentes funções que vão desde a avaliação formativa informal até à avaliação sumativa formal. Entre elas, encontram-se a avaliação formativa formal e a avaliação sumativa informal que diferem pelo facto da primeira utilizar a informação recolhida para adaptar o ensino que se desenvolve na sala de aula, enquanto a segunda utiliza essa informação para verificar os progressos dos alunos face ao que foi anteriormente estabelecido. Já Fernandes (2008), não compreende esta perspectiva com facilidade, porque as duas têm propósitos distintos,

ocorrem em momentos distintos e têm uma inserção pedagógica distinta. Para o autor são complementares. As relações entre ambas permitem que a informação recolhida para fins formativos possa ser usada para fins sumativos e vice-versa. Defende que uma avaliação sumativa interna de qualidade deve basear-se nos princípios, nos métodos e nos conteúdos da avaliação formativa alternativa.

Por vezes as diferentes funções da avaliação estão em tensão, o que leva alguns a defenderem que um só sistema de avaliação não serve todas as funções e são necessários sistemas separados (Wiliam, 2004). Para o autor, esta separação excluiria os professores das avaliações sumativas, ou então operariam em paralelo, mas em distintos sistemas para as funções formativas e sumativas, o que resultaria sempre na marginalização da função formativa. Sistemas integrados podem servir as várias funções e a função formativa deve ser a base da avaliação.

Actualmente, as políticas educativas investem mais na avaliação sumativa estando os procedimentos de avaliação melhor estabelecidos que na avaliação formativa (Fernandes, 2005). A avaliação sumativa comanda de tal ordem as preocupações de todos que é significativo o facto de haver estudos em Portugal que revelam que as práticas de avaliação formativa parecem estar longe da vida das escolas (Fernandes, 2009).

No entanto, convém sublinhar que a própria Organisation for Economic Co-operation and Development [OECD] (2008) refere a propósito das avaliações externas que, não há nada inerente às avaliações sumativas que impeçam os professores de realizar avaliações formativas. A informação de natureza sumativa pode ser usada formativamente. Podem-se identificar forças e fraquezas e decidir sobre estratégias de melhoria.

De facto, a base de qualquer sistema de avaliação deve estar assente na avaliação para fins formativos. Mesmo quando as avaliações são por natureza sumativas, devem assumir uma função formativa, procurando melhorar as pessoas, os seus projectos e as suas organizações.

Na perspectiva de Wiliam (2004), a avaliação é formativa para os indivíduos quando podem usar o feedback da avaliação para melhorar a sua aprendizagem. É formativa para os professores quando os resultados da avaliação, adequadamente interpretados, ajudam a melhorar o seu ensino, em tópicos específicos ou na generalidade. É formativa para escolas e regiões se a informação recolhida puder ser interpretada de maneira a melhorar a qualidade das aprendizagens entre escolas e regiões. É formativa para o sistema quando for usada para melhorar o desempenho do próprio sistema.

Assim, concluímos de acordo com Bonniol (2001a) que, quando as prioridades institucionais passarem da selecção e do controlo para a formação e as aprendizagens essenciais, a avaliação formativa, integrará muito bem a avaliação sumativa.

### **1.1.2. Qualidades formais da avaliação**

Nas actividades relacionadas com a avaliação incluem-se as situações de prova. Popham (2003) considera que as informações que os professores recolhem sobre o desempenho dos alunos são geradas nessas situações que são tentativas formais<sup>3</sup> para avaliar conhecimentos, habilidades e atitudes.

Por outro lado, Mertler e Campbell (2005), apoiando-se noutros autores, estimam que os professores passam mais de 50% do seu tempo em actividades relacionadas com avaliação, mas sentem-se muito mal preparados para enfrentar esse desafio.

Todavia, Fernandes (2006) considera que a avaliação não é uma questão de fé ou convicção, apesar de não ser matéria exacta, tem de estar suportada em evidências que sejam consistentes e tenham significado. Na sua síntese da bibliografia, também Mertler e Campbell (2005) sublinham que a informação gerada pela avaliação deve ser precisa e ter significado, deve ser válida e fiável. Popham (2003) refere, mesmo, que a literacia em avaliação é impossível sem se compreender minimamente o que são conceitos como a validade, a fiabilidade e os enviesamentos, considerando que deles dependem as inferências em que se pode confiar. Para este autor, inferências inválidas podem conduzir a conclusões inválidas de pais, administradores e políticos sobre aquilo que os professores e as suas escolas estão a fazer.

Neste quadro, consideramos que a avaliação das aprendizagens não é alheia a valores que representam qualidades formais da avaliação como validade, fidedignidade, objectividade, ou aos seus equivalentes dos modelos qualitativos, a saber: credibilidade/transferabilidade, confiabilidade e confirmabilidade (Coryn, 2007).

Como a teoria ainda não adequou os conceitos às perspectivas actuais da avaliação, eles devem ser entendidos mais como orientações do que como regras absolutas.

Assim sendo, neste ponto vamos discutir as qualidades formais que qualquer avaliação deve respeitar e que, segundo Harlen (2007), são critério para se poderem considerar e apreciar os prós e contras de cada tipo de avaliação. Discutimos o critério da validade de constructo e da fiabilidade e discutimos a relação estreita que deve existir entre as duas para que as inferências que se fazem possam ser dignas de confiança.

---

<sup>3</sup> Para este autor o adjectivo formal faz a distinção das apreciações informais que os professores fazem no dia a dia.

### *Validade*

A validade é o grau de adequação entre o que se declara fazer, como avaliar esta ou aquela dimensão, e aquilo que efectivamente se faz, entre o que o instrumento mede e aquilo que pretende medir (De Ketele & Gerard, 2005).

Portanto, na definição clássica de validade, uma prova válida é, aquela que avalia, realmente, aquilo que pretende avaliar (Marion et al., 2002). Para estes autores, vários esforços para responder à questão de saber se uma prova avalia aquilo para que foi construída seguiram vários caminhos, ao longo dos anos, e que correspondem aos vários tipos de validade<sup>4</sup>.

Há, no entanto, um acordo geral que diferentes tipos de validade como a validade concorrente ou a validade de conteúdo estão contidos no conceito global de validade de constructo (Gipps, 1994). Segundo Marion et al. (2002), a compreensão sobre o conceito de validade e que diz respeito à validade de constructo é, actualmente, a visão standard. Portanto, o constructo<sup>5</sup> é o tema unificador do conceito unitário que é a validade (Gipps, 1994).

Harlen (2007) define a validade de constructo em termos de o que é avaliado corresponder bem ao comportamento ou resultado de aprendizagem que se pretende avaliar. Assim, é necessária uma clarificação do domínio a ser avaliado. Para Marion et al. (2002), a validade de uma prova refere-se ao processo de definição do constructo, ou seja, a característica ou conceito que se expressa na acção ou desempenho.

Para vários autores, este tipo de validade depende da utilização que se faz do instrumento e dos resultados, podendo ser válida para determinadas utilizações e inválida para outras. A validade de constructo não diz respeito só à prova em si mesma, ou aos resultados que produz, mas depende das inferências que se façam e das decisões que se tomem (Gipps, 1994; Marion et al., 2002; Harlen & Gardner, 2010; Stobart, 2010).

---

<sup>4</sup> Fernandes (2005), revendo a literatura, refere esses diferentes tipos de validade, a saber:

- Validade de previsão – bom indicador de desempenhos futuros.
- Validade de conteúdo – se contem amostra significativa do conteúdo do domínio a avaliar.
- Validade concorrente – se resultados se correlacionam com outros sobre as mesmas aprendizagens. Em contexto de avaliação padronizada, o método correlacional é o mais frequentemente utilizado para verificar a validade. Se a correlação entre o novo instrumento e um amplamente aceite que avalie o mesmo constructo for alta, a evidência é suportada (Moon, Brighton, Callahan & Robison, 2005).
- Validade de critério – se prever o desempenho relativo a determinado critério.
- Validade de constructo – se for medida adequada do constructo, ou seja, da competência subjacente que está a ser avaliada.

<sup>5</sup> Constructo - domínio ou competência a ser avaliada (Stobart, 2010).

Nesta linha, Popham (2003) explica que não é a prova em si que pode ser válida ou inválida, mas antes as inferências que se fazem baseadas no desempenho dos alunos. Para este autor a pergunta que se deve fazer é se a inferência baseada na classificação, pontuação ou notação, é válida, considerando que a análise de validade deve centrar-se nas inferências próprias da prova, ou seja, aquelas que a prova permite realizar. Se a interpretação sobre o significado das classificações for precisa pode dizer-se que se chegou a uma inferência válida baseada na prova.

Uma outra forma de utilizar e interpretar as informações geradas pela avaliação ganhou progressivamente importância e é conhecida pela validade consequencial (Meisels, Liaw, Dorfman & Nelson, 1995). Gulikers et al. (2004), na leitura que fazem da matéria publicada, explicam que a validade consequencial descreve os efeitos intencionais e não intencionais da avaliação no ensino e na aprendizagem. Para Meisels et al. (1995) há consequências positivas, se a avaliação permitir melhorar a instrução ou o controlo dos alunos sobre a sua própria aprendizagem, ou há consequências negativas, se a avaliação levar ao estreitamento do currículo ou à redução da autonomia do professor.

#### *Fiabilidade*

A fiabilidade é o grau de confiança que se pode ter nos resultados observados. Implica serem os mesmos se a recolha de informação for noutro momento, feita por um outro instrumento, ou por uma outra pessoa. Um ponto de convergência desejado independente de avaliadores e de circunstâncias (De Ketele & Gerard, 2005).

Portanto, o exemplo clássico de verificação da fiabilidade é representado pela possibilidade de um aluno obter o mesmo resultado se uma prova for repetida (Marion et al., 2002). A fiabilidade é, assim, frequentemente definida como a medida em que a avaliação, se repetida, dá o mesmo resultado e refere-se ao facto dos resultados terem uma consistência<sup>6</sup> aceitável para uma utilização específica (Harlen, 2007).

Explicando melhor, no nosso entendimento a fiabilidade verifica-se se, o mesmo aluno, nas mesmas condições ou contexto, com as mesmas capacidades ou prontidão, realiza a mesma prova da mesma maneira - consistência do desempenho. Mas, no tempo que separa duas repetições é praticamente impossível estarmos perante a ausência de

---

<sup>6</sup> Popham (2003) define três tipos de consistência:

- «Stability» – a estabilidade diz respeito à consistência com que uma prova avalia alguma coisa independentemente do momento em que é aplicada. Quer dizer, se não houver instrução, realizada a prova, agora ou passado um mês, obtém-se o mesmo resultado.
- «Alternate-form» – havendo duas formas alternativas da mesma prova, que sejam equivalentes, se alguém as realizar, os resultados forem muito similares.
- «Internal consistency» – existe consistência interna quando todos os itens da prova têm o mesmo poder de discriminação.

Segundo este autor uma prova pode ter grande consistência interna e não ter estabilidade.

treino ou processo de maturação. Assim, faz mais sentido falar no registo do desempenho do aluno avaliado obter, repetidamente, a mesma classificação (pontuação ou notação). Trata-se, então, de consistência na classificação, ou seja, de objectividade do avaliador.

No entanto, é difícil alcançar a objectividade na avaliação e tem de se evitar que a subjectividade, que até pode ser uma componente essencial do acto de avaliar, se transforme em arbitrariedade e ausência de critérios (Gerard, 2006; Serpa, 2010).

De facto, pode não haver consistência, se os resultados forem influenciados por quem conduz as avaliações (Harlen, 2007). É o que ocorre quando as correcções variam com correctores diferentes, principalmente, em questões não objectivas de resposta aberta (Kellaghan & Madaus, 2003).

“Comme le démontrent la plupart des traités de docimologie, les juges, totalement abandonnés à eux-mêmes, sont rarement d'accord” (V. De Landsheere, 1988, p. 136). A autora refere que, por vezes, podem observar-se situações dramáticas a propósito do mesmo desempenho, com alunos que numa turma passam e com o mesmo desempenho reprovam noutra. Explica, também, que na educação, os juízos de valor têm sempre a última palavra, sendo que todos os métodos procuram o máximo de objectividade possível, que é definida como a concordância relativa dos juízos dos especialistas.

Assim, é importante, por um lado, conhecer a extensão do acordo e do desacordo entre avaliadores (Moon et al., 2005), e por outro lado, treinar os avaliadores até que estejam alcançados níveis elevados de acordo (Barret, 2001; Eckes, 2005).

#### *Relações entre validade e fiabilidade*

De acordo com Marion et al. (2002), “Both validity and reliability are important, but validity is more important” (p. 23). Para estes autores a fiabilidade deve ser vista como um dos componentes da validade e consideram que quando ela falha a validade enfrenta muitos obstáculos, mas sublinham que, a fiabilidade sendo condição necessária, não é, no entanto, suficiente para que haja validade.

Para Harlen (2007; 2004), a validade e a fiabilidade não são independentes, existindo uma relação estreita entre elas. Essa relação expressa-se frequentemente de uma forma que torna a fiabilidade o principal requisito. Assim, quando se força a fiabilidade pode minar-se a validade de constructo, desde que se opte por itens de prova que permitam obter resultados cuja avaliação seja mais fiável. Ao reduzir a validade, o conhecimento e as competências avaliados passam a não reflectir o que se pretende.

O principal problema que reflecte este efeito de favorecer a fiabilidade, comprometendo a validade, consiste em estreitar a extensão do conteúdo e as competências a avaliar (Marion et al., 2002)



É o que ocorre quando apenas se inclui conhecimento factual, com respostas que podem ser marcadas inequivocamente e se exclui o que é mais difícil de julgar, como a aplicação de conhecimento, o raciocínio crítico (Harlen, 2007; 2004) e as produções autênticas, ou seja, o desempenho em situações problemáticas representativas do constructo.

Numa perspectiva que relaciona as qualidades formais da avaliação com os seus custos, Stobart (2010) considera que a validade é frequentemente limitada quando se tentam conseguir avaliações com grande fiabilidade e sem custos. Por outro lado, também considera que procurar grande validade de constructo a custos baixos pode determinar pouca fiabilidade e que assegurar, em simultâneo, grande validade de constructo e grande fiabilidade, pode implicar uma avaliação muito onerosa.

Em investigação, os critérios de validade, fidelidade e objectividade estão muito ancorados na tradição quantitativa e quando combinados têm normalmente o estatuto de «trindade» (Coryn, 2007). Mas, nas ciências humanas são mitos, quer a avaliação totalmente objectiva, quer a validade garantida sem condições. Por isso, independentemente daquilo que se faça, qualquer julgamento pode ser sempre contestado (V. De Landsheere, 1988).

De facto, todas as avaliações estão sujeitas ao erro (Harlen, 2004). O erro sendo inerente à condição humana pode também ser um instrumento útil para o seu próprio progresso (Fragata & Martins, 2004). Para Bonniol e Vial (2001), nunca conseguiremos neutralizar todos os factores e fontes de erro, considerando que nem as situações e ferramentas de avaliação serão imaculadas, nem o avaliador irá libertar-se da sua condição humana. Também Clímaco (2005), sintetizando o pensamento de outros autores, refere que uma dos aspectos da avaliação, qualquer que seja a abordagem, é a existência duma projecção dos valores do avaliador no modo como observa a realidade.

Referindo-se às produções autênticas mais padronizadas na forma e, ao mesmo tempo, obrigatoriamente originais no conteúdo – as teses de doutoramento, Coryn (2007) afirma que para tomar em conta as realidades da pesquisa qualitativa e a complexidade dos fenómenos sociais, há que redefinir-se os termos, sublinhando que os critérios que têm prevalecido até agora devem ser mais utilizados como guias do que como uma ortodoxia.

De acordo com Coryn (2007), a tradição qualitativa está agora a desenvolver a sua própria «trindade»<sup>7</sup>, ou seja, a utilizar novos critérios.

Talvez na linha de Coryn, Gipps (1994) afirma que é fundamental reconceptualizar a fiabilidade para que a teoria da avaliação educativa se desenvolva, até porque cada vez mais se avaliam tarefas complexas em contextos muito diversos. Para Fernandes (2006) precisamos de outros conceitos mais em sintonia com a avaliação formativa alternativa ou outra maneira de pensar a validade e a fiabilidade.

Neste quadro, partindo do pressuposto que os conceitos discutidos são apenas guias de orientação nos processos de avaliação, optámos por considerar que a forma de Stobart (2010) e Harlen (2007) pensarem as qualidades formais da avaliação e os seus custos, corresponde a um entendimento das relações de dependência entre elas que representa uma boa base para se discutir a relação dessas qualidades formais com os diferentes tipos de avaliação.

Concluímos então que, em primeiro lugar, há que garantir a validade de constructo. Não nos interessam avaliações que apesar de consistentes não representem o que os alunos sabem e são capazes de fazer. Essas avaliações consistentemente erradas levam a inferências que não são válidas. Em segundo lugar, trata-se de assegurar a máxima fiabilidade, de uma forma tal, que seja possível preservar a validade de constructo a custos aceitáveis.

Por outro lado, dependendo do tipo de avaliação, a relação de dependência entre validade, fiabilidade e custos deve variar. É o que vamos passar a discutir no ponto que se segue.

*As qualidades formais da avaliação na Avaliação Interna:*

*Avaliação Formativa*

Na avaliação formativa as aprendizagens que desejamos que ocorram são a grande referência para apreciarmos a validade de constructo e, segundo Harlen (2007) a sua interpretação está relacionada com a progressão nessas aprendizagens. O que se pretende globalmente da avaliação formativa é que haja um impacto positivo na aprendizagem e no ensino. A fiabilidade não deve merecer preocupação se a avaliação for

---

<sup>7</sup> Para este autor, na pesquisa qualitativa ganham espaço novos critérios, tais como: a fidedignidade que integra por um lado, a credibilidade, ou seja, a dimensão em que os resultados são credíveis e, por outro, a transferabilidade, isto é o grau em que os resultados podem ser generalizáveis noutro contexto; a confiabilidade, que representa possíveis mudanças de contexto dentro de cada pesquisa, até porque cada investigador tem uma perspectiva única para o estudo; a confirmabilidade que se refere ao grau em que os resultados podem ser verificados por outros. Numa curiosa comparação com os critérios quantitativos, este autor considera os subcritérios credibilidade e transferabilidade congruentes respectivamente com a validade interna e a validade externa, o critério confiabilidade congruente com a fiabilidade e o critério confirmabilidade congruente com a objectividade.

cuidadosa, se a informação recolhida tiver muito valor para apoiar a aprendizagem (Harlen, 2007). Os custos desta avaliação praticamente não existem, os recursos materiais e o tempo que ela requer são quase irrelevantes, correspondem apenas à preparação e planeamento feitos por cada professor.

Harlen (2007) refere que a especificidade e o detalhe dos critérios com que se julga a validade de constructo na avaliação formativa devem ser maiores que na avaliação sumativa, podendo aproveitar-se as acções necessárias para melhorar a fiabilidade da informação para a avaliação sumativa, de forma a influenciarem a fiabilidade das apreciações feitas com objectivos formativos. Para esta autora só acontecerão efeitos negativos, se os professores seguirem procedimentos mecânicos.

#### *Avaliação Sumativa realizada ao nível da Turma*

A validade de constructo será tanto maior quanto a informação recolhida reflectir as aprendizagens que se querem avaliar (Harlen, 2007), ou seja, se houver uma cobertura adequada dos objectivos prioritários e que devem corresponder ao que foi ensinado. Para a autora, a relação de dependência entre a validade e a fiabilidade deve pender a favor da validade. De facto, de pouco servem grandes níveis de fiabilidade se as inferências que se fizerem não poderem ser consideradas válidas. De qualquer forma, cada professor deve minimizar as possíveis ameaças à fiabilidade (Harlen, 2007), procurando que as formas de avaliação que utiliza lhe garantam consistência nas apreciações. O impacto da avaliação será sempre positivo se as suas características, eminentemente sumativas, integrarem as grandes prioridades que devem ser a formação e as aprendizagens.

Harlen (2007) afirma que o que é ensinado é o que deve ser testado, e sublinha que, utilizar testes desenvolvidos externamente fornece menos informação com grande validade de constructo para objectivos relacionados com a avaliação interna que as avaliações desenvolvidas pelos professores, já que a relação não é tão estreita com os objectivos de aprendizagem de um dado período.

Por outro lado, V. De Landsheere (1988) refere que, uma prova em que todas as questões façam parte de um domínio fechado, tem automaticamente validade de conteúdo, sendo que nos casos representativos de um domínio aberto devem procurar-se indicadores fortes.

Para Popham (2003), os professores podem verificar se um determinado item reflecte suficientemente o conteúdo que supostamente representa, utilizando painéis de validação com outros professores e pais, sem ter de desenvolver um enorme esforço.

Este autor também chama à atenção para o facto da validade das inferências poderem ser diminuídas por factores externos como por exemplo um aluno de trigonometria com paralisia cerebral que não controla bem um lápis para desenhar linhas curvas, um

estrangeiro recém-chegado que não domina a língua, ou alguém que dormiu pouco e não consegue completar toda a prova.

Quanto à fiabilidade, Popham (2003) refere que não é necessário qualquer tipo de evidência, já que o benefício seria demasiado pouco, em comparação com o esforço requerido.

Harlen (2007) considera que a forma de minimizar as ameaças à fiabilidade e aumentar a fiabilidade dos julgamentos feitos pelos professores é envolvê-los em tarefas conjuntas que permitam alinhar os seus juízos.

A mesma autora considera que estes procedimentos podem ajudar os professores a moderar os seus julgamentos, embora essa necessária moderação, que pode assegurar fiabilidade, requer, por um lado, tempo de escola, mas por outro lado, custa menos que a moderação entre escolas quando a avaliação sumativa é externa.

Quando há um processo de moderação dos resultados, não só a fiabilidade das avaliações dos professores é comparável à dos exames, como o próprio processo de moderação representa uma oportunidade valiosa de aprendizagem profissional (Harlen & Gardner, 2010).

Concluimos que o principal custo desta avaliação é o tempo que o professor despende na elaboração, administração e correcção de testes ou na concepção de outras tarefas de avaliação e respectiva recolha de informação e interpretação do desempenho dos alunos.

#### *Provas de aferição realizadas ao nível do Grupo/Departamento*

Como argumenta Fernandes (2005), a propósito das provas de aferição realizadas a nível nacional, e que consideramos ser válido para as provas de aferição realizadas ao nível do Grupo/Departamento de cada escola, estas provas podem avaliar domínios do currículo que não podem ser testados pelos exames; podem ser realizadas em disciplinas que não são testados pelos exames; podem dar sinal aos alunos e professores do que é necessário valorizar; podem ser administradas nas salas de aula, o que lhes confere uma natureza menos formal que os exames, e o tempo e os procedimentos podem ser flexibilizados e menos padronizados.

Por outro lado, estas provas permitem monitorizar áreas particulares e fazer comparações de ano para ano. Se, por exemplo, por ano, disciplina, género, domínios dos programas, os dados que se recolhem, forem agregados, ficaremos com um retrato fiel do que é a escola. Como referem Rink et al. (2002), saber como as escolas vão indo é mais importante que saber como cada aluno vai.

No entanto, o currículo deve estar de tal forma clarificado para todos os professores do Grupo/Departamento de forma a haver a garantia que todos os alunos tiveram a mesma

oportunidade de aprendizagem e a prova reflecte o principal do currículo. Para que isso aconteça, é necessário tempo de trabalho em conjunto. Tempo que também é necessário para aferir critérios de avaliação que garantam que os avaliadores apreciam o mesmo desempenho de maneira igual.

Mas, estas provas também podem fazer aumentar a confiança colectiva, no assumir de responsabilidades perante a comunidade, sobre o que se faz em cada escola. Gallagher e Ratzlaff (2008), consideram que os dados que a avaliação gera podem ser usados para objectivos de prestação de contas, se uma avaliação válida e fiável estiver incorporada na instrução e no currículo, em vez de se impor de fora da escola, formatando as práticas pedagógicas. Os autores afirmam que as decisões mais importantes sobre ensino e aprendizagem acontecem nas salas de aula e é aí, portanto, que se deve assegurar a validade e fiabilidade da avaliação.

Segundo V. De Landsheere (1988), a responsabilidade de prestar contas sobre a qualidade do trabalho que se desenvolve, um dos princípios fundamentais da democracia, corresponde a um direito dos cidadãos em relação às instituições públicas. Para Clímaco (2005), quem assume esta responsabilidade “é credível porque mostra que sabe o que faz, explica como se faz e porquê” (p. 60).

Concluimos que estas provas, não tendo grande tradição em Portugal, podem ter um impacto muito positivo nas escolas, quer para professores, quer para alunos. Podem representar a oportunidade de influenciar positivamente o ensino e a aprendizagem se, aquilo que é essencial no currículo de cada disciplina for testado, e se não se basearem só, nas disciplinas que habitualmente têm exames. Por um lado, a validade das inferências que se possam fazer dos resultados destas provas deve ser grande, e por isso, devem maioritariamente ser baseadas no desempenho, envolvendo de preferência tarefas autênticas. Por outro lado, não se justificam grandes preocupações com a fiabilidade já que estas provas não têm impacto directo na progressão ou na certificação de cada aluno.

*As qualidades formais da avaliação na Avaliação Externa:*

*Avaliação Sumativa – Os Exames*

Os exames certificam as aprendizagens realizadas durante um ano ou ciclo de escolaridade, conferindo, frequentemente, um diploma que permite ter acesso a um emprego ou a ser seleccionado para entrar numa escola do ensino superior. Nalguns sistemas de ensino os resultados também são usados para a avaliação do desempenho docente ou para comparar escolas.

Para Harlen (2007), os exames necessitam de grande validade e de grande fiabilidade. Esta autora refere o requisito de validade, ao considerar que esta avaliação manda uma grande mensagem sobre o que é valorizado na aprendizagem. Quanto à

fiabilidade considera que só quando ela é grande se pode confiar nos resultados e garantir que a classificação alcançada significa o mesmo para todos os alunos independentemente de quem tenha feito o julgamento, sublinhando que, sem esta precisão, os alunos podem ser afectados se quiserem entrar em escolas selectivas ou cursos do ensino superior.

A mesma autora, ressalva que, se os exames só estreitarem a gama de objectivos que podem ser avaliados com maior fiabilidade, a avaliação não dará a informação que é necessária, mas, se cobrir toda a gama de objectivos e ajudar a clarificar-lhes o significado o impacto da avaliação pode ser positivo nas experiências de aprendizagem. Pelo contrário o impacto pode ser negativo se houver um desencontro entre currículo e avaliação, sendo mais negativo se os resultados forem usados para avaliar professores e escolas. Também defende que quando a informação é recolhida em avaliações externas, tipo exame, isso põe pressão nos professores que começam a ensinar para passar no exame e, como os resultados significam menos aprendizagem e mais habilidade para responder às questões dos exames, a validade é infringida.

Os exames têm sido só aplicados a algumas disciplinas consideradas importantes no currículo, com destaque para os resultados cognitivos (Clímaco, 2005), centram-se nos conteúdos, e privilegiam o conhecimento dos conteúdos e não a sua mobilização para resolver problemas (Fernandes, 2005).

A maior limitação é que o tipo de avaliação padronizada que habitualmente se utiliza é muito restrita, tendo como prioridade avaliar conhecimentos memorizáveis e habilidades analíticas relacionadas com o conhecimento, ficando quase sempre de fora a criatividade, o pensamento prático, a sabedoria (Sternberg, 2008).

O mesmo tende a acontecer relativamente aos «resultados sociais de educação», como a atitude face à escola, o comportamento ou o desenvolvimento moral e cívico, que são os mais difíceis de avaliar (Clímaco, 2005), se forem seguidos os formatos analíticos dos testes padronizados.

De acordo com Harlen e Gardner (2010), são poucas as vezes que aparecem nos exames e testes desse tipo, rubricas ou questões a requerer aos alunos que sejam criativos e apresentem argumentos, ou que demonstrem compreensão de situações e temas complexos. Mesmo quando tais rubricas são incluídas nas provas de avaliação externa, a pressão que os professores sentem, leva-os a ensinar de maneira demasiado circunscrita às tarefas de preparação dos exames em vez de ajudar os alunos a compreender o que está a ser avaliado. Em consequência há alunos que nunca aprenderam os conceitos fundamentais que estão implicados nessas mesmas provas.

Por outro lado, é frequente que os responsáveis pelas avaliações em larga escala atribuam grande ênfase à fiabilidade e à eficiência, destacando o dinheiro gasto e o tempo

ocupado, precisamente, à custa da validade de constructo (Stobart, 2010). O autor considera que este problema é exemplarmente demonstrado pela limitada representação do constructo num exame de Línguas, ao reduzir-se a prova a tarefas de leitura e de escrita, negligenciando assim as competências orais, de expressão e compreensão, que são essenciais na aprendizagem de uma língua.

No entanto, parece que nem sempre as preocupações com a fiabilidade se traduzem no rigor que os seus defensores preconizam. Focando o caso português, Fernandes (2005) afirma que a baixa fiabilidade entre os correctores nos exames nacionais é preocupante, o que remete para um maior cuidado na elaboração dos critérios de correcção e para a necessidade de existência de acções de formação e de processos de aferição de critérios.

Concluimos que os exames têm um forte impacto na vida dos alunos e dos professores e representam muito tempo gasto na sua preparação, realização e classificação. Como os seus resultados necessitam de ser considerados credíveis consideramos normal exigir-se maior rigor no processo de exames do que na avaliação sumativa interna. No entanto, não se poderão considerar credíveis os resultados que, embora fiáveis, não representem o desempenho dos alunos nos domínios e competências consideradas mais importantes nos *currícula*.

#### *As Provas de Aferição*

As considerações feitas sobre a validade, fiabilidade e o impacto no ensino e na aprendizagem, a propósito das provas de aferição realizadas a nível de Grupo/Departamento de cada escola, parecem-nos válidas, e são as mesmas, para as provas de aferição realizadas a nível nacional. Fernandes (2005), analisando criticamente, a prática anterior em provas de aferição, refere vantagens que considera boas, se as provas fossem pensadas de outra forma, a saber:

- Provas de natureza mais aberta e diversificada. Como a progressão não está em causa, a fiabilidade não se põe.
- Flexibilização dos procedimentos de administração que a aproximem a contextos de sala de aula.
- Testar temas curriculares não cobertos pelos exames.

Por outro lado, na perspectiva de Harlen e Gardner (2010), como a preocupação não está centrada nos resultados individuais, mas no desempenho da população escolar, em cada domínio de aprendizagem, não há necessidade de todos os alunos responderem aos mesmos itens. Deste modo, podem usar-se um maior número de itens para representar uma boa amostra do domínio e, assim, garantir o aumento da validade da prova.

As diferenças principais das provas de aferição nacionais, relativamente às provas de aferição na avaliação interna, são as seguintes: as provas de aferição a nível nacional permitem obter um retrato que não é apenas local mas que pode reflectir o país todo, o que é positivo; podem ter uma maior influência no ensino e na aprendizagem porque sinalizam fortemente as competências mais valorizadas; implicam mais recursos pelas exigências de tempo e dinheiro.

No entanto, como refere Fernandes (2005), podem ser uma forma mais barata de recolher informação, induzir boas práticas, apoiar decisões políticas, e promover a investigação.

#### *As Provas dos estudos internacionais sobre avaliação*

As provas dos estudos internacionais sobre avaliação, como o PISA e os da IEA, têm o mérito de procurar avaliar o uso que os alunos fazem do conhecimento, em itens sobre tarefas do dia-a-dia, e não só o que eles sabem acerca dos conteúdos. Aparentemente, este facto permite garantir alguma validade nas inferências que se fazem dos resultados dessas provas. Mas, como refere Fernandes (2005, 2007), os sistemas educativos são muito diferentes, fazem parte de países que culturalmente e socialmente são diferentes, que têm expectativas diferentes do que se espera dum jovem de 15 anos, que é a idade dos alunos que são testados no PISA, tendo como consequência que a relevância curricular dos assuntos que integram os itens não é a mesma para todos os países. Também Crahay (2002), a propósito da IEA, se refere ao facto de haver grande dificuldade em construir um teste internacional que leve em conta as especificidades dos programas de cada país participante, ou seja, que tenha validade de conteúdo.

Mas, também, a consistência dos resultados parece oferecer dúvidas. Para Fernandes (2005, 2007), os países participantes podem não seguir os procedimentos metodológicos definidos pela coordenação dos estudos, sendo muito difícil comparar resultados, o que a torna numa comparação que carece de consistência. No entanto, para Afonso e Costa (2009), quem toma decisões políticas está ávido de estudos internacionais, assistindo-se hoje a uma substituição do debate mais ideológico por processos de decisão política que se baseiam em dados e números. Para os autores, pretende-se que a informação técnica e o conhecimento técnico legitimem a política educativa.

Para Fernandes (2005, 2007), tal como nas provas de aferição, parece que só quem participa nesses estudos, conhece os seus objectivos, as metodologias, o que se investiga, e as suas próprias limitações. Para este autor, por um lado, é difícil definir uma metodologia que avalie os resultados e, por outro lado, é necessário contextualizar as análises que se fazem dos resultados para que as conclusões sejam mais rigorosas e tenham fundamento. Por exemplo, em Portugal os alunos de 15 anos estão distribuídos por



vários anos de escolaridade e não têm prática de resolução de testes de escolha múltipla, o que representa uma desvantagem. Considera, também, que os relatórios deveriam ser mais interpretativos que descritivos. No entanto, conclui que o caminho não está em rejeitar a utilidade dos estudos argumentando com as suas limitações actuais ou na sua aceitação como indicadores do fracasso das escolas.

*Em conclusão*

Analizadas e discutidas as funções e as qualidades formais da avaliação na sua relação com os diferentes tipos de avaliação, parece-nos pertinente argumentar que em termos ideais, quer na avaliação interna, quer na avaliação externa, as diferentes provas de avaliação devem ter uma matriz comum, ou seja, a estrutura da prova, as situações ou tarefas de avaliação que avaliam os vários domínios e os respectivos critérios de avaliação devem ser muito similares. Consideramos que uma mesma matriz pode dar resposta aos objectivos de um exame ou de uma prova sumativa na avaliação interna.

Se integradas nesta perspectiva, as provas de aferição a nível nacional podiam dar um contributo notável para testar o compromisso que é necessário para garantir nas diferentes avaliações a validade de constructo, a máxima fiabilidade possível e custos razoáveis.

Também na avaliação para fins formativos estas considerações nos parecem válidas. Se forem utilizadas as mesmas tarefas de avaliação, se elas se confundirem com as experiências de aprendizagem, é dada, aos alunos, a oportunidade de praticar o que se quer que aprendam e que vai ser avaliado. É claro que as tarefas de avaliação não devem representar a actividade exclusiva nas salas de aula. É desejável que os alunos também pratiquem outras tarefas que facilitam a aprendizagem.

Assim, no quadro da EF, uma situação de avaliação nos JDC que avalia o desempenho dos jogadores no jogo, deve ser a mesma numa prova de aferição a nível nacional, numa prova de avaliação interna, na avaliação inicial durante a 1ª etapa do ano lectivo, ou nas avaliações que os professores realizam com fins formativos. Como vimos anteriormente, apenas deve variar a exigência com os níveis de fiabilidade e com os custos da avaliação. Naturalmente, que uma prova que certifica as aprendizagens exige uma fiabilidade maior, e também é maior o tempo gasto na preparação de uma prova de aferição que numa avaliação para fins formativos.

## **1.2 Problemas tradicionalmente associados à Avaliação Escolar**

Vários problemas que estão identificados e estudados têm permanecido ao longo dos anos na realidade das escolas. Sendo actuais, têm, nalguns casos, ganho praticamente o estatuto de problemas tradicionais. Se não, vejamos:

A avaliação está pouco relacionada com o processo de ensino aprendizagem e orienta-se principalmente para classificar os alunos (Fernandes, 2005; 2007). Nalguns casos, os resultados de final de ano são tão valorizados que as decisões sobre progressão ou retenção dos alunos e a avaliação da qualidade do trabalho das escolas dependem deles (Tucker & Stronge, 2005; Barton, 2008; Popham, 2003).

Há a convicção que a consistência da avaliação interna é débil (Fernandes, 2005; 2007), o que não ajuda o esforço de auto-avaliação de muitas escolas, continuando a acreditar-se que os exames, normalmente centrados nos conteúdos (Fernandes, 2005; Roldão, 2003), melhoram as aprendizagens (Fernandes, 2007) e são o verdadeiro critério para avaliar o desempenho das escolas (Darling-Hammond, 2004; Popham, 2003; Guskey, 2005).

Dá-se muita importância a habilidades discretas e analíticas (Nazario, 2005; Sternberg, 2008; McTighe & O'Connor, 2005), não se mobiliza conhecimentos para resolver problemas (Fernandes, 2005) e não se encoraja a colaboração (Herrington & Herrington, 2006). Os modelos que ainda prevalecem na avaliação e no ensino são responsáveis, em parte, pela falta de competências dos alunos (Roldão, 2003), pelo abandono precoce e também pelas elevadas taxas de retenção (Fernandes, 2007).

Analisemos, então, estes problemas com mais pormenor, discutindo algumas das suas implicações.

### *Avaliação e classificação*

De acordo com Fernandes (2005), a avaliação confundida com medição tem, ainda hoje, uma grande influência na maioria dos sistemas de educação. É uma avaliação descontextualizada, com muito pouca participação dos alunos e que se centra nos conhecimentos. Para este autor, somar, classificar e seleccionar são as suas grandes funções, utilizando-se provas de avaliação para quantificar aprendizagens e aptidões que são comparadas e ordenadas em escalas. Estas escalas de medida foram precisamente criadas para resolver o problema do valor (Bonniol & Vial, 2001). No entanto, “uma parte fundamental da educação só é mensurável através de medidas indirectas ou então de medidas qualitativas” (Clímaco, 2005, p. 90).

A abordagem tradicional de avaliação consiste principalmente em testes descontextualizados, com itens desenhados psicometricamente para testar conhecimentos e habilidades cognitivas de nível inferior, sendo usados de forma sumativa para diferenciar

alunos e colocá-los em ranking (Gulikers et al., 2004). Como na avaliação psicométrica os instrumentos são referidos à norma, os alunos ao não controlarem as classificações dos colegas ficam impedidos de controlar as suas próprias (Fernandes, 2005).

Para Fernandes (2005), há progressos nas posteriores gerações de avaliação, a avaliação como descrição e a avaliação como juízo de valor, que já consideram a avaliação de professores, de projectos, de currículos e de políticas, ainda assim, apresentando muitas limitações porque pouco consideram os contextos e a diversidade de culturas e valores, só admitem falhas de alunos e professores e estão, também, muito dependentes do paradigma positivista em que quem avalia não pode contaminar os resultados.

Em Portugal, a nota e aquilo que ela permite obter ainda dominam, na maior parte dos casos, as preocupações e as práticas dos professores. Avaliar torna-se então sinónimo de classificar e pouca atenção é dada ao processo que conduz ao produto. Para Fernandes (2005, 2007), predomina uma avaliação, essencialmente, para classificar, em detrimento de uma avaliação que permita melhores aprendizagens, sendo que a avaliação sumativa está muito associada ao progresso académico e à obtenção de diploma e, por isso, o seu impacto social é grande.

Ainda se ensina para depois avaliar e os momentos de avaliação são pontuais e têm, normalmente, um impacto elevado numa classificação que nem sempre traduz a capacidade de utilizar os conhecimentos, as técnicas e as habilidades em situações de vida real. Roldão (2003) considera que sistemas como o nosso, em que há uma dramatização da classificação, não são exigentes no que se refere à necessidade de comprovar se a nota corresponde a saberes reais. Quem tem boas classificações pode nunca aprender a usar o que sabe (Roldão, 2005).

Mesmo só pensando em termos de nota/classificação é redutor aceitar que a média de momentos pontuais de avaliação traduzem toda a complexidade das operações necessárias à construção de um conhecimento que, num mundo moderno, se pretende eficazmente utilizado. De acordo com Pacheco (1994), a forma como os professores operacionalizam os instrumentos de avaliação revela uma sobrevalorização dos testes sumativos e muito pouco de uma recolha de dados estruturada e sistematizada que represente uma alternativa ao que é tradicional.

De acordo com Roldão (2003), esses testes ou fichas cobrem conteúdos desenvolvidos num determinado período de tempo e é raro avaliarem o que os alunos sabem fazer com eles. Para a autora, não basta conhecer e memorizar conteúdos é preciso demonstrar que são dominados e que se sabem usar, sendo frequente os bons alunos revelarem-se pouco competentes, quando têm de aplicar conhecimentos ou resolver

problemas concretos, e também, revelarem dificuldades quando colocados fora do formato tradicional de teste ou exames.

Fernandes (2005) considera que o problema não está nos testes mas na utilização que se possa fazer deles e dos seus resultados e sublinha que o processo de atribuição de classificações é, em si mesmo, complexo. Para este autor, um passo a dar seria relativizar o peso dos testes sumativos. Nesta linha, também Roldão (2003) refere que a complexidade não se pode reduzir a uma média da nota dos testes, que antes se chamavam pontos e agora se chamam fichas, e considera que essas fichas têm tido muito pouco de formativo, já que de pouco têm servido para explicar e remediar dificuldades surgidas.

Os momentos de demonstração de resultados exigem mais. Exigem que se rompa com a tradição dos testes de papel e lápis a cobrir matérias assimiladas de memória e se utilizem formas alternativas de avaliação que possam estar alinhadas com o currículo e conduzam o processo de ensino aprendizagem.

#### *Avaliação e instrução*

Na nossa experiência pessoal nos Conselhos de Turma, um dos indicadores mais claros da pouca relação que ainda existe entre a avaliação e a instrução é o predomínio de práticas de planeamento que não tomam em consideração o ponto de partida dos alunos no momento em que se inicia um ano lectivo.

“The way a teacher designs tests and applies test data can profoundly affect how well that teacher teaches” (Popham, 2003, p. 1). De acordo com este autor, a avaliação pode permitir um conhecimento prévio do aluno, pode permitir determinar quanto tempo leva a ensinar algo e pode determinar a eficácia da instrução. Explica, também, que diferente avaliação faz apelo a diferente actividade de instrução

Segundo Porcher (2001), normalmente só é dada atenção ao ponto de chegada e muito raramente se toma em consideração o ponto de partida. Para o autor é importante elaborar ferramentas de avaliação que permitam definir o ponto de partida dos alunos.

Mesmo quando se encontram práticas de avaliação que antecedem a instrução, elas são marcadamente diagnósticas. Só permitem constatar factos. Os testes ou fichas parceladas em segmentos de matéria mais pequenos (Roldão, 2003), feitos numa sessão sobre determinado domínio do programa, utilizam-se muito. Como contêm o vício da exclusividade do que é pontual ou momentâneo, apenas permitem inferir o que os alunos sabem ou não sabem. Pouco ou nada permitem saber sobre se aprendem rápido ou não, se os bons alunos são individualistas ou, pelo contrário podem ser utilizados como catalisadores das aprendizagens dos mais fracos, ou ainda, se as características que se observam aconselham mais determinada formação de grupos ou outra, e em que momentos.

Na verdade, se o plano anual da turma estiver todo absolutamente elaborado antes de conhecer os alunos, isso, quererá dizer que as dinâmicas que se estabelecem, as relações entre alunos, as suas lideranças, a forma como aprendem, as suas potencialidades e limitações não foram suficientemente tomadas em consideração, porque se desconhecem.

*Resultados de final de ano*

Nos sistemas de avaliação escolar os resultados de final de ano são muito valorizados e tem grande impacto, quer nas decisões de passagem ou retenção dos alunos, quer nas apreciações que se fazem ao trabalho dos professores e à qualidade das escolas.

Os exames assumem um carácter importante na apreciação que se faz sobre se os alunos tiveram ou não sucesso nas suas aprendizagens. Para além de elemento regulador e certificador das aprendizagens têm servido, também, para comparar escolas.

Revendo a literatura, Tucker e Stronge (2005) consideram que os resultados em provas de aproveitamento absoluto tendem a expressar mais aquilo que os alunos trazem para a escola, e menos aquilo que eles aprendem na escola. Estes autores destacam o facto de prevalecer a noção de que na escola, a aptidão é que conta e não a progressão ou a melhoria.

Barton (2008) considera que este sistema não mede o progresso de nenhum aluno, turma ou escola. Para o autor, os testes ou provas de avaliação tipicamente realizadas no final de cada ano lectivo medem o conhecimento absoluto do aluno, não interessando quando ou onde o adquiriram, se na mesma escola ou noutras escolas em anos anteriores, na família, ou durante o verão.

É precisamente, por reflectirem o efeito acumulado daquilo que aprenderam em casa e na escola nos anos anteriores que, de acordo com Tucker e Stronge (2005), “as notas ou os resultados dos testes de avaliação dificilmente poderão ser consideradas medidas válidas indicadoras da influência do professor durante um determinado ano” (p. 17).

Fazer comparações de resultados nos finais de anos é falacioso porque os grupos de alunos em cada turma também mudam (Popham, 2003), e, para além disso, essas comparações não tomam em consideração que o grupo de alunos de um ano lectivo é diferente, na sua preparação educativa, do grupo de alunos de outro ano lectivo (Barton, 2008).

De facto, considerar os resultados de exames ou de outras avaliações de final de ano muito importantes é não só compreensível como aceitável. Considerá-los em exclusivo, sem tomar em consideração os pontos de partida, é que já é abusivo e pode traduzir-se em deduções equivocadas e sem validade.

Por outro lado, Barton (2008) defende que um sistema só é transparente e compreensível se tomar em consideração as perdas ou ganhos típicos do verão.

Na verdade, o Verão representa para muitos alunos uma oportunidade acrescida para aprender e para outros alunos uma ausência absoluta de estimulação que se possa considerar significativa. Uns participam em campos de férias, campos de treinos de modalidades desportivas, em viagens com significado histórico e cultural relevantes, em oficinas de escrita, teatro ou dança, etc., outros têm um tempo de paragem de três meses porque só na escola encontram essa oportunidade de aprendizagem.

Se ficar claro o que são resultados escolares, se eles forem entendidos como os progressos reais nas aprendizagens, provavelmente muitas das apreciações enviesadas não ocorreriam. Para Tucker e Stronge (2005), válido é “comparar os ganhos de aprendizagem estabelecendo a ligação entre dois momentos temporais diferentes para os mesmos alunos e não para grupos diferentes de alunos” (p. 105).

Para Clímaco (2005), “reconhece-se, generalizadamente, que existem grandes limitações na utilização dos resultados dos alunos como medidas da qualidade das escolas ou como equivalentes de “avaliação educativa”, devido ao carácter restrito deste modo de traduzir a noção de “resultado” e de “eficácia” (p. 130).

Apesar desse reconhecimento persistem as comparações entre escolas que não tomam em consideração os seus pontos de partida. É por isso, que de acordo com Darling-Hammond (2004) e Fernandes (2005), as escolas são estrategicamente muito defensivas para obterem bons lugares nos rankings. Segundo estes autores, quando os resultados das provas de exame se podem traduzir em recompensas ou sanções, empurram os alunos fracos para a educação especial, convidam alguns a desistir, tentam que os piores não façam exame ou que anulem matriculas para que esses resultados possam parecer melhores.

Para além disto, Wainer (2007), sublinha que a ignorância sobre a «equação de Moivre»<sup>8</sup> tem sido responsável por muitos equívocos ocorridos nos Estados Unidos. Pensava-se que o facto de alguns estados terem percentagens baixas de cancro do rim tinha relação com estilos de vida rurais. Exemplos semelhantes aconteceram a propósito das cidades mais seguras ou das escolas com melhores resultados em provas nacionais. De facto, o que têm em comum os estados com mais alta e mais baixa percentagem de cancro do rim, as cidades mais e menos seguras e as escolas com os melhores e os piores resultados, é a pouca população.

Na verdade, como a variação da média é inversamente proporcional ao tamanho da amostra, as escolas mais pequenas têm maior variação que as escolas grandes. Isto significa que estarem colocadas nos primeiros lugares ou nos últimos lugares dos rankings

---

<sup>8</sup> Segundo o autor, Moivre mostrou em 1730 que o desvio padrão não aumenta na proporção da amostra. Em vez disso, é proporcional à raiz quadrada do tamanho da amostra.

pode, apenas, dever-se a esta simples «equação de Moivre» e não a quaisquer características de desempenho.

Assim, podemos argumentar que se o objectivo principal fosse tornar cada um melhor e a escola melhor e a avaliação fosse vista como uma necessidade de todos e não uma exigência de poucos, talvez estas tendências não ocorressem. Se essas comparações tomassem em consideração as características da comunidade, o tipo de população alvo, os recursos disponíveis, mas sobretudo os pontos de partida, em particular dos alunos, os rankings teriam uma legitimidade que agora não possuem. Como refere Fernandes (2005), monitorizar o desempenho das escolas não tem de passar pela publicação de rankings, até porque podem surgir tantos rankings como os que quisermos dependendo dos critérios adoptados.

#### *Consistência da avaliação interna e mecanismos de autoavaliação das escolas*

“No ensino básico, há um currículo nacional orientado por competências. Há razões para crer que a consistência da avaliação interna é débil e que, sendo assim, podemos pôr em causa a equidade, pois não sabemos se os alunos têm todos acesso ao mesmo tipo de oportunidades de aprendizagem. Parece que ninguém estará em condições de garantir que, ao longo do país, todos os alunos tiveram oportunidade para aprender este ou aquele domínio essencial e estruturante do currículo” (Fernandes, 2005, p. 138).

A convicção na debilidade da avaliação interna, manifestada por este autor, é consistente com a forma como analisamos a realidade escolar e com a opinião de Harlen (2007) que afirma ser amplamente aceite que a fiabilidade dos julgamentos dos professores é baixa. De facto, observa-se que no quadro concreto de cada escola, onde os professores têm as condições mais privilegiadas para avaliar os seus alunos, são muito diversificados os critérios e instrumentos de avaliação utilizados o que torna muito difícil a interpretação dos resultados.

“Há também quem critique o peso relativo da avaliação interna, considerando-o excessivo, argumentando que a validade e a fiabilidade das avaliações realizadas nas salas de aula são geralmente fracas” (Fernandes, 2007, p. 593).

Vale a pena argumentar se não se trata de um problema de formação, o facto de os professores usarem instrumentos formativos que não têm a fiabilidade necessária, para fazer a avaliação sumativa. Inclusivamente, há professores que utilizam as avaliações que os alunos esperam ser apenas formativas para tomar decisões de certificação (De Ketele & Gerard, 2005).

Por outro lado, têm faltado instrumentos de avaliação sumativa em larga escala que possam servir de referência para os professores. Na verdade, pode dizer-se que os exames e as provas de aferição são recentes e a própria experiência de construção colectiva de instrumentos de avaliação na avaliação interna, como era o caso das provas globais,

praticamente acabou, sendo poucas as escolas que as utilizam como instrumento de aferição de conhecimentos.

Mas, também, as escolas necessitam de mecanismos de auto-avaliação que sejam catalisadores da mudança (Pacheco, 1994). A auto-avaliação pode constituir-se como o processo de aprendizagem da própria escola (Clímaco, 2005). Ora, esta debilidade na avaliação interna também não ajuda ao esforço de auto-avaliação que muitas escolas procuram efectuar. À difícil interpretação de resultados, devido à enorme variedade de critérios e instrumentos de avaliação, junta-se o pouco hábito de interpretação orientado para tomadas de decisão. Ficamo-nos por uma recolha exaustiva de dados sobre classificações, assiduidade, anulações, apoios concedidos, etc., e olhamo-los sem procurar ler e agir em conformidade. É aquilo que, Clímaco (2005) considera ser uma produção estatística sem avaliação. “Na medida em que não “lêem” os resultados, não se apercebem do que é preciso fazer para os alterar, nem do que precisam saber para mudar ou melhorar a situação” (p. 233).

#### *Provas de avaliação externa e monitorização*

De acordo com Popham (2003), a principal razão para o aumento do número de provas de exame é que quem decide politicamente quer evidências fortes de qual o desempenho das escolas públicas. Para este autor, não só os políticos mas, também, a maioria dos cidadãos acreditam que os resultados dos exames é o verdadeiro critério pelo qual se pode medir a eficácia das escolas, imperando a lógica que considera boas as escolas com melhores resultados e, más as que têm piores resultados e sublinha que, deste modo, os professores são postos sob pressão no que respeita a melhorar os resultados dos alunos em provas de exame. Para Darling-Hammond (2004), as provas de exame são informação importante para um sistema de prestação de contas, mas não são o próprio sistema.

Na verdade, a investigação ainda não provou que se melhoram as aprendizagens, aumentando o número de exames, mas os governos de vários países continuam a insistir neste rumo (Fernandes, 2007). Não é por um aluno responder às perguntas de um exame que ele vai utilizar esses conhecimentos na vida real (Cardinet, 2001a). No entanto, os membros das administrações têm grande confiança na validade das provas de exame padronizadas à norma, sendo muitas vezes essas avaliações em larga escala utilizadas em exclusivo para tomar decisões de grande impacto sobre as escolas, professores e alunos (Guskey, 2005; Darling-Hammond, 2004).

Quando se implementam testes padronizados de grande impacto podem verificar-se muitas das armadilhas que lhes estão associados, como estreitamento do currículo, maior ênfase na memorização do que em competências de nível superior, e distorção dos



próprios resultados (Gallagher & Ratzlaff, 2008). Discutindo os trabalhos publicados sobre o problema, Darling-Hammond (2004) afirma que várias pesquisas têm mostrado essas consequências não intencionais nas provas de exame, como estreitamento do currículo e conduzir a instrução na direcção de habilidades cognitivas de nível inferior.

Popham (2003) refere que é comum, os professores passarem muito tempo a fazer testes com itens, praticamente, iguais aos que saem nos exames. Na perspectiva de Stobart (2010), isto acontece porque alunos querem ter boas classificações e as escolas querem causar boa impressão com os seus resultados.

Fazendo o ponto da discussão teórica, Nazario (2005) reforça o facto deste tipo de ensino poder dar uma imagem inflacionada do desempenho, porque os alunos podem parecer dominar o que é ensinado, sem terem, de facto, compreendidos os conceitos subjacentes. Uma resposta correcta pode corresponder apenas a um hábito adquirido em exercícios muito repetidos (Cardinet, 2001b). Para V. De Landsheere (1988), representa uma aprendizagem forçada pelas circunstâncias e que se esquece de seguida porque não está assente em motivação profunda. Para a autora isto é consequência dos alunos estudarem em função de se ter ensinado o tipo de desempenho que a prova exige.

De acordo com Popham (2003), há diversos tipos de provas que fazem inclinar os professores a fazerem só isto. Para este autor, a aposta da instrução não deveria ser na preparação para as provas, mas sim na preparação para as habilidades, conhecimentos e atitudes que elas representam.

Por outro lado, segundo McMillan (2005), a literatura que relata a influência de provas de grande impacto, nomeadamente as que implicam decisões de progressão e retenção para os alunos ou sanções e recompensas para as escolas e os professores, mostra que não existem apenas consequências negativas, mas que se misturam os impactos positivos e negativos. Fernandes (2005) destaca algumas vantagens dos exames, como dar indicações acerca do que é importante ensinar e aprender, induzir novas práticas ao nível do ensino e da avaliação, moderar as avaliações internas, ajudar as escolas a melhorar os seus projectos, e contribuir para avaliar o sistema educativo ajudando a que se tomem decisões a todos os níveis.

Greene, Winters & Forster (2004) mostraram que as provas de grande impacto não distorcem a informação sobre desempenho dos alunos e que os seus resultados podem ser usados como ferramentas fiáveis de política. Discutindo os resultados do seu estudo, os autores sugerem que, distorções como, ensinar para a prova, fazer batota ou outro tipo de manipulações, não produzem resultados muito diferentes, dos resultados das provas que não têm grande impacto na vida dos alunos e das escolas.

McMillan (2005) nas conclusões da sua pesquisa, sugere que houve efeitos positivos das provas de exame, quer na instrução e quer na própria avaliação, verificando-se maior incidência em processos cognitivos de nível superior, por contraste com a memorização, e também maior ênfase na avaliação formativa. Contudo, refere que a principal limitação destas conclusões está no facto dos dados serem fruto de auto relatórios dos professores e que, portanto, as mudanças ocorridas poderem ser muito diferentes das relatadas.

Num escrutínio aos principais argumentos que os críticos das provas padronizadas utilizam, Phelps (1999) relembra que, o que determina o carácter do processo cognitivo necessário para responder correctamente não está no formato da resposta mas na estrutura da pergunta.

Para Roldão (2003), a questão que se coloca não é a do desaparecimento das avaliações externas, o que se deve questionar é a natureza dessas provas. Para a autora, a nossa prática curricular, que é exaustiva, leva a que se questione mais o domínio de conteúdos do que as competências que, através deles, se manifestam.

Essa tradição centralista também ocorre mesmo nos bons exemplos como, por vezes têm sido, as provas de aferição em Língua Portuguesa e Matemática. Para Roldão (2005), estas provas têm procurado afastar-se das questões que apenas procuram avaliar o grau de conformidade com os conteúdos. No entanto, a demora exagerada na chegada dos relatórios de análise dos resultados às escolas (Fernandes, 2007), tem impedido uma interpretação que permita agir com base nos resultados em tempo útil. É, precisamente, esse retorno de informação que permitirá, de acordo com V. De Landsheere (1988), que os professores se perguntem porque os resultados dos seus alunos diferem dos seus colegas. Para Fernandes (2005), era importante discutir e analisar os resultados das avaliações externas e verificar quais os seus efeitos na escola, através da comparação entre o currículo que é avaliado na avaliação externa e o que é avaliado na avaliação interna.

Por outro lado, a avaliação pode tornar evidente muitas das nossas deficiências. Pode permitir que se revejam práticas educativas. Como destaca (Fernandes, 1994):

“Importa contribuir para que a avaliação externa (aferida) venha a constituir um instrumento de avaliação do currículo, da qualidade de ensino e da aprendizagem dos estudantes assim como um veículo que pode induzir, nas escolas, novas formas de ensinar e de avaliar” (p. 25).

No entanto, essa indução de novas práticas só será proveitosa se não se traduzir numa aplicação mecanicista. Se não houver a necessária reflexão, quando a referência for boa copiar-se-ão as formas, mas também o mesmo ocorrerá se a referência for má. De acordo com Fernandes (2005):

“Não se tem feito a pedagogia necessária para se entender o alcance que estas provas podem ter na melhoria da qualidade da educação básica, na formação de

professores, no desenvolvimento de bancos de itens a utilizar pelos professores e pelas escolas ou no desenvolvimento da investigação numa área em que estamos francamente atrasados em relação aos nossos parceiros europeus” (p. 124).

Então, não basta termos exames para que o processo de monitorização seja eficaz. É necessário outro tipo de exames, centrados na avaliação de competências, é necessária prática reflexiva que a partir da interpretação de resultados, conduza a tomadas de decisão de aperfeiçoamento do sistema e é necessária investigação que apoie e sustente as práticas de avaliação. Como refere Fernandes (2005), a avaliação certificativa também pode ser útil e contextualizada, diversificada e inteligente. Não é razoável pensar que o ensino e a avaliação vão melhorar só porque há exames. Se os professores se obrigarem a atingir determinados resultados com todos os alunos, será evidente a utilidade dos dispositivos de pilotagem do rendimento escolar e da avaliação externa para fins de regulação (Crahay, 2002).

Muito se tem discutido a utilidade da existência de provas de avaliação externa no nosso sistema de ensino. Não nos parece que a extinção pura e simples dos exames ou das provas de aferição resolva algum problema de fundo. Estão ainda por aprofundar as questões que poderão introduzir mudanças numa prática que é profundamente centralista e que não promove a reflexão necessária sobre os benefícios que se poderiam alcançar na melhoria da qualidade da educação.

*Currículo, instrução e os problemas da vida real*

“When I look at the skills and concepts I have needed to succeed in my own field, I find a number that are crucial: creativity, common sense, wisdom, ethics, dedication, honesty, teamwork, hard work, knowing how to win and how to lose, a sense of fair play, and lifelong learning. But memorizing books is certainly not one of them” (Sternberg, 2008, p. 20).

Na linha de Sternberg, também, V. De Landsheere (1988) considera como habilidades decisivas para a nossa época, a criatividade, a curiosidade de espírito, a capacidade de trabalhar em equipa,

No entanto, não é por se adquirirem determinadas capacidades, conhecimentos ou habilidades na escola que elas serão bem utilizadas na vida real, embora, segundo V. De Landsheere (1988), quando em função dos resultados de uma avaliação se estima que um aluno é competente, pensa-se que ele é capaz de aplicar as aquisições em estudos posteriores e em situações de vida real. Para Popham (2003), um aluno que tenha competências generalizáveis, está capaz de aplicá-las em situações dentro e fora da escola.

Trata-se, então, de trazer para o essencial do currículo aprendizagens que se possam generalizar e formas de participação dos alunos que lhes garantam autonomia e responsabilidade. De acordo com Fernandes (2005), a forma tradicional de aprender está longe de corresponder já que a resolução de problemas e de situações relacionadas com a

vida real que potenciariam o desenvolvimento de processos mais complexos de pensamento, estão ausentes. Para este autor, a própria avaliação do trabalho também não ajuda a que os alunos aprendam melhor e com mais compreensão e não os inclui como participantes activos no processo de avaliação e desenvolvimento das suas aprendizagens.

Como destaca Roldão (2005), é necessário aprender a usar o saber. Para a autora, a apropriação dos conceitos, saberes, técnicas, factos e teorias demonstra-se através da utilização contextualizada do que se aprendeu,

O problema é que esta avaliação não mede tudo aquilo que é necessário para ter sucesso na vida do dia-a-dia (Sternberg, 2008).

“Too often, classroom instruction and assessment overemphasize decontextualized drills and provide too few opportunities for students to actually «play the game»” (McTighe & O’Connor, 2005, p. 11). Para este autor, isto representa o mesmo que um jogador só treinar cantos e nunca jogar, ou um nadador só «fazer piscinas» e nunca competir em provas. Na sua leitura da matéria publicada, Nazario (2005) reforça este aspecto referindo que as maiores críticas aos testes padronizados são a de eles só terem uma resposta certa, estreitarem o currículo, focarem-se em competências discretas, não medirem competências cognitivas de ordem superior, prestarem mais atenção ao produto que ao processo e não considerarem a criatividade.

Herrington e Herrington (2006) argumentam que nas universidades continua a transmitir-se conhecimento teórico em fragmentos analíticos para alunos passivos receberem e consumirem, estuda-se por sebatas ou notas de leitura, decorar para os exames é a estratégia mais comum de aprendizagem e a colaboração não é solicitada ou encorajada. Para estes autores, este é o resultado de, também assim, muitos professores terem sido ensinados. Também Fernandes (2005, 2007), considera que continua a ensinar-se procedimentos rotineiros e que praticamente só se reproduz informação que foi previamente transmitida, faltando dar significado ao que se aprende e relacionar com aprendizagens anteriores.

Pacheco (1994) considera que estamos perante a “ausência de uma teoria da aprendizagem suficientemente integrativa das dimensões cognitiva, afectiva e motora” (p. 124). De facto, ainda imperam práticas que fraccionam os domínios, quer para a formulação de objectivos de aprendizagem, quer para a avaliação da consecução desses mesmos objectivos, como se pode constatar a partir dos critérios de avaliação aprovados na maioria das escolas. Na verdade, o comportamento humano é um todo e ao não se tomar isso como ponto de partida, mais longe se estará de um currículo que possa dar resposta aos grandes desafios da educação do nosso tempo.

Mas, também, há uma diferença entre o que se ensina e o que se pensa que os alunos aprendem (V. De Landsheere, 1988). Como ainda se acredita que a fala dos professores produz conhecimento nos alunos (Roldão, 2003), as metodologias de apresentação estão assentes no discurso de um professor ou no manual (Roldão, 2000). O manual é muitas vezes confundido com o programa, substituindo-se a ele e ao próprio professor. Trata-se de uma tendência que representa, para Pacheco (1994), uma limitação na possibilidade de o professor desempenhar o papel de um verdadeiro agente curricular. Por seu turno, Hansen e Stephens (2000) consideram que as palestras são o principal método de ensino, não sendo mais que uma transferência de informação de uma só via.

Outro problema no ensino é a abordagem pedagógica que pretende tratar os alunos da mesma maneira (Roldão, 2000), embora pelas suas características, aptidões e necessidade eles sejam muitos diferentes, ao nível da turma, do ano de curso ou da escola. É um tratamento que produz mais desigualdade (Crahay, 2002), onde a indiferença à diferença se traduz na aplicação de um ensino massivo para toda a turma.

Para Roldão (2001), a escola está organizada em torno dos princípios da homogeneidade, segmentação, sequencialidade e conformidade e precisamos de princípios que enquadrem um novo paradigma, em que se passe a utilizar a heterogeneidade como centro do trabalho escolar. Para a autora, no paradigma que vigora, as respostas organizativas ainda são típicas do período anterior à massificação em que o conjunto dos alunos era mais homogéneo e a lógica predominantemente selectiva. Uma selecção que dantes era anterior à entrada no ensino e que hoje se faz no interior da própria instituição por processos escolares (Crahay, 2002).

Mas a lógica disciplinar existente também dificulta a possibilidade do currículo interagir mais com as problemáticas da vida real. De acordo com Clímaco (2005) e Roldão (2001), as estruturas profundas das escolas não se alteraram, sendo que a turma continua como unidade na organização dos alunos e da escola, e os currículos estão fragmentados numa lógica que é sobretudo disciplinar e individual na forma de encarar o trabalho docente. Para Roldão (2000), há uma confusão entre disciplina curricular e disciplina científica. “A dominância da lógica disciplinar tem sido frequentemente confundida, nos debates curriculares, com a questão epistemológica disciplinaridade/ interdisciplinaridade dos saberes científicos” (p. 9). A autora considera que o aspecto central nesta questão é o modo de organização do trabalho e não a natureza do saber.

Ora aqui está justamente o problema específico do ensino e da avaliação dos Jogos Desportivos Colectivos, e de outras matérias da Educação Física, em que o “trabalho de grupo” decorre simultaneamente, por um lado, do processo, do modo de organização e

também, por outro lado, das características do saber em foco, ou seja as características da própria actividade referente.

A colaboração entre professores de diferentes campos de saber e a resolução de problemas que solicitem a integração de saberes está muito dificultada. A carga institucional que é dada aos conselhos de turma é muito inferior relativamente aos conselhos de agrupamento — veja-se, por exemplo, as diferenças no número de representantes no conselho pedagógico (Roldão, 2000). A autora sugere que há uma dificuldade de instituir espaços e tempos de trabalho colaborativo nos conselhos de turma, fracassando as tentativas de coordenação entre disciplinas.

Assim, não espanta que os professores valorizem, quase em exclusivo, os conhecimentos das disciplinas que leccionam. De acordo com Fernandes (2005), é esta questão cultural que leva os professores a organizar a comunicação na sala de aula em torno de conteúdos específicos e a não definir competências de natureza transversal como alvos a atingir. É verdade que não se pode pensar sem conteúdo (Sternberg, 2008), mas os conteúdos não podem ser fins em si mesmos.

Para Roldão (2003), os conteúdos “são indispensáveis para ganhar alguma coisa que se não tinha antes, para nos tornarmos mais competentes — cientificamente, linguisticamente, historicamente, esteticamente, matematicamente” (p. 16).

Seguimos a argumentação de Pacheco (1994), segundo o qual o currículo escolar deve ser um conjunto integrado pelo princípio da interdisciplinaridade e não apenas um somatório das várias disciplinas. Nesta linha, o autor defende que só com uma nova cultura de colaboração se poderá ultrapassar o individualismo docente, caracterizado por cada aula ter fronteiras disciplinares intransponíveis.

#### *Taxas de reprovação e falta de competências*

De acordo com Fernandes (2007), os principais problemas endémicos com que nos debatemos são a utilização exclusiva da avaliação para classificar, os fracos resultados dos alunos em provas de avaliação externa e as elevadas taxas de abandono e retenção. Para este autor, um insucesso tão expressivo põe em causa o desenvolvimento do país e tem custos económicos e ao nível da coesão social. Ora, paradoxalmente, o sistema de avaliação tem soluções e processos que deveriam contrariar isto, como o predomínio da avaliação formativa e da avaliação interna e a regra segundo a qual a decisão de retenção é tomada no final do ciclo de escolaridade.

A retenção pode ser uma falsa resposta à questão das dificuldades e ritmos diferenciados de aprendizagem. De facto, Darling-Hammond (2004) afirma estar demonstrado que reter os alunos não os ajuda a ter sucesso na escola. A autora refere que se olharmos para as práticas nas salas de aula, vemos que muito pouco é feito para garantir

uma experiência de qualidade e mais adequada às necessidades, simplesmente, repete-se de novo a mesma experiência.

Está descrita na literatura (Crahay, 2002) uma tendência em que, perante forte heterogeneidade entre as capacidades dos alunos, os mesmos professores que facilmente reprovam os fracos são renitentes em fazer avançar um ano os alunos brilhantes. De acordo com este autor, os estudos apoiados pela IEA revelaram que os países altamente selectivos no ensino secundário não produzem uma elite de mais alto nível quando comparados com os que optaram por democratizar.

Ao contrário de abandono e retenção, Pacheco (1994) preconiza a ideia de ajuda, em que se oferecem oportunidades de recuperação com apoios e complementos, educativos. No entanto, ao contrário do que seria de esperar muitas provas de avaliação não são mais do que constatações de incapacidade (V. De Landsheere, 1988).

É muito criticada a escola pública pelos lugares que ocupa nos estudos comparados internacionais (Clímaco, 2005). Os alunos evidenciam problemas no desenvolvimento de competências superiores de pensamento, sendo fraco o desempenho na resolução de problemas, na aplicação e utilização do conhecimento em novas situações, ou na análise e interpretação da informação (Fernandes, 2007). Mas, também, a escola pública é criticada por haver alunos que a deixam sem qualificações suficientes e com falta de competências básicas para entrar no mercado de trabalho (Clímaco, 2005).

Para Roldão (2003), é muito grave que alunos abandonem a escola com escassas competências, apesar de terem dado muitas matérias. Para a autora, todos precisam de aprender porque a sociedade exige mais qualificação, sublinhando que a falta de integração social dos que não aprenderam o que é proposto nos *curricula*, facilita o aumento de grupos de marginais e excluídos.

Concluimos que a falta de competências básicas revelada em largos sectores da população escolar, o abandono da escolaridade obrigatória que ainda é muito expressivo, e as reprovações que sistematicamente ocorrem em todo o sistema de ensino, são resultado de um ensino e avaliação que revelam problemas recorrentes. Como refere Fernandes (2007), as elevadas taxas de abandono e retenção estão, certamente, relacionadas com as formas que se utilizam para avaliar e ensinar.

### **1.3. A Avaliação Autêntica no quadro da Avaliação Alternativa**

Nos dias de hoje, num mundo em mudança permanente e acelerada, as tendências da sociedade moderna colocam novos e exigentes desafios à escola. Em sintonia com um novo perfil de cidadania e de vida activa reclama-se por um ensino e avaliação diferentes que seja alternativa ao que tem sido tradicional em educação. Segundo Sternberg (2008), a sociedade precisa de pessoas sábias, criativas e práticas. Para este autor, aqueles que só

memorizam, que se centram em factos e figuras, não estarão tão preparados para um mundo que muda muito rapidamente, ao nível político, económico, social e, até mesmo, ao nível do contexto climático.

São mudanças que implicam adaptações a um mundo necessariamente incerto. Assim, de acordo com Gulikers et al. (2004), a educação actual deve centrar-se no desenvolvimento de alunos competentes e não tanto em simples aquisição de conhecimentos, pretendendo-se que o conhecimento seja construído a partir duma aprendizagem que se quer activa e reflexiva e com procedimentos de avaliação contextualizados e baseados no desempenho. Como destaca Pacheco (1994), a escola deve ser um local de construção de conhecimento e onde, com a ajuda do professor, o aluno constrói a sua aprendizagem a partir de metodologias activas centradas na resolução de problemas.

Actualmente, a sociedade já não se contenta com qualidades que eram consideradas suficientes no passado. Políticos e empregadores clamam por outras qualificações: “What employers, governments and nations require are graduates that display attributes necessary for knowledge building communities: graduates who can create, innovate, and communicate in their chosen profession” (Herrington & Herrington, 2006, p.2).

Por outro lado, Sternberg (2008) considera que apesar da dificuldade de ensinar e avaliar, o aspecto mais importante e o mais negligenciado na educação de hoje é, precisamente, a capacidade de usar o conhecimento sabiamente, com empenhamento ético. Nesta linha, também, Crahay (2002), argumenta que é necessária nova ética para pensar a escola, uma ética que respeite a diversidade cultural e promova a solidariedade.

Estamos, então, perante novas exigências curriculares que colocam, aos estudantes, desafios mais ambiciosos que a simples memorização. De acordo com Fernandes (2005) a escola actual deve preocupar-se com aspectos como a resolução de problemas, a utilização das novas tecnologias, a comunicação, o exercício da crítica, o trabalho em grupos multiculturais e a resolução de situações complexas. O mesmo autor refere, ainda, a pressão que os estudos internacionais sobre avaliação exercem sobre os vários países participantes no sentido dos currículos e da avaliação estarem alinhados com essas novas exigências. Também Caldwell (2000), revendo a literatura, refere a criatividade, resolução de problemas, trabalho em equipa, literacia e numeracia como os aspectos que o currículo deve promover.

Nesta mesma linha, Harlen e Gardner (2010) consideram como elementos essenciais da educação do século 21, a resolução de problemas, o pensamento crítico, o empreendedorismo e a cidadania. Na sua perspectiva o conhecimento compreendido é aquele que é usado a resolver problemas ou a tomar decisões e implica o desenvolvimento



da capacidade de aplicar a aprendizagem em contextos diferentes daqueles em que foram aprendidos. Para os autores uma avaliação válida deve estar alinhada com estas preocupações.

Em contraponto com as convicções e as práticas que estão na origem dos problemas associados à avaliação escolar, e em alternativa às formas mais tradicionais de avaliação, novas formas de avaliação tentam responder às actuais exigências curriculares. De entre as diferentes avaliações alternativas, a chamada avaliação autêntica (Wiggins, 1989; 1990) tem-se revelado muito promissora no sentido da melhoria das práticas e do conceito de avaliação.

Deste modo, vamos em primeiro lugar, analisar e discutir as características principais que devem ser assumidas na avaliação alternativa, em segundo lugar, analisar as especificidades da avaliação autêntica no quadro da avaliação alternativa, discutindo as suas vantagens e os problemas que lhe podem estar associados e, por último, discutir a necessidade de uma nova cultura de escola para que a avaliação autêntica ganhe espaço.

#### **1.3.1. Características principais na Avaliação Alternativa**

A) Uma Avaliação que vise principalmente processos complexos de pensamento e competências do pensamento de nível superior.

“The goal of assessment is the acquisition of higher-order thinking processes and competencies instead of factual knowledge and basic skills” (Gulikers et al., 2004, p. 67).

As competências são saberes em utilização (Fernandes, 2005), são construídas sobre conhecimentos e experiências, mas vão muito para além deles (Roldão, 2003). Quando se quer resolver um problema ou assumir uma tarefa, é competente quem tem a capacidade de mobilizar conhecimentos conceptuais e de procedimento (Crahay, 2002).

A competência é, portanto, da ordem do «saber mobilizar», e não só simples aplicação (Roldão, 2003; 2005). Aliás, a autora considera que qualquer situação de avaliação tem que permitir demonstrar as competências que estiverem em causa. Nesta perspectiva, seja qual for a forma organizativa utilizada, como um exame, uma questão, uma tarefa, uma actividade ou um jogo, é na situação de avaliação que se expressa a competência (Roldão, 2003).

É, portanto, em contexto que os saberes se mobilizam e a competência se torna visível, não dependendo do contexto em que foi adquirida porque é na capacidade de uso em novos contextos que ela se institui (Roldão, 2005). Como refere Fernandes (2005), os conhecimentos devem ser mobilizados para utilizar em contextos diversificados.

Assim, quando se quer preparar os alunos para transferir situações da sala de aula para situações reais, devem ser colocados perante um problema a resolver em situações mobilizadoras, suficientemente complexas e realistas (Cardinet, 2001b). Importa fornecer

um contexto autêntico que reflecta a maneira como o conhecimento será usado na vida real. Isso implica que as tarefas sejam complexas e contenham actividades com relevância no mundo real (Herrington & Herrington, 2006).

Uma abordagem baseada em competências implica avaliar os conhecimentos através de situações complexas que necessitem da parte do aluno uma produção, ela própria complexa, em que mobilizam um conjunto integrado de recursos para resolver a situação (De Ketele & Gerard, 2005).

Por outro lado, nem tudo na vida é complexo e os objectivos de aprendizagem apontam para diferentes direcções que a avaliação também deve contemplar. “Assessment is critical to determine whether or not change is taking place and, if so, if the change is moving in a desirable direction” (Nazario, 2005, p. 3). Revendo a bibliografia, Gulikers et al. (2004) defendem que, para os alunos serem competentes a lidar com problemas profissionais, a tarefa deve ter semelhanças com as características da situação na vida real que é o seu referente. No entanto, para os autores, a vida nem sempre apresenta situações complexas para resolver e o próprio exercício da memória também é importante. Sublinham o facto de na vida real haver sempre problemas simples que se podem estruturar na avaliação com soluções simples e sem necessitar de interdisciplinaridade.

Nazario (2005) é um autor, entre outros, que defende que os testes tradicionais de resposta simples, se forem cuidadosamente construídos e utilizados, podem ser uma ferramenta muito útil porque ainda são necessários para memorizar factos, compreender cálculo matemático básico ou identificar pontos de uma narrativa.

Todavia, importa considerar o facto de Sternberg (2008) afirmar que, mesmo quando os alunos são apenas avaliados sobre aspectos relacionados com a memória, eles evidenciam um melhor desempenho quando são ensinados de uma forma ampla do que quando são ensinados só para memorizar.

B) Uma Avaliação centrada nas estratégias metacognitivas<sup>9</sup> que os alunos utilizam e como processo negociado e participado.

A regulação externa que orienta a auto-regulação do aluno não o pode substituir. Pode regular-se os procedimentos, métodos e técnicas utilizados, mas jamais regularemos directamente os processos internos do aluno (Bonniol & Vial, 2001). Herrington e Herrington (2006) são autores, entre muitos outros, representativos da ideia que a interacção do professor com os alunos deve ocorrer principalmente a um nível metacognitivo. Também, na linha de outros autores, Gulikers et al. (2004), especificando o conceito de competência,

---

<sup>9</sup> «Meta cognição» implica consciência de como se vai na aprendizagem. Aluno que tem consciência de como aprende é mais capaz de atingir metas, desenvolver várias estratégias de aprendizagem, e controlar e avaliar o seu próprio processo de aprendizagem (OECD, 2008).

sublinham a importância dos alunos desenvolverem competências metacognitivas como a reflexão e, não só, competências cognitivas como resolução de problemas ou pensamento crítico.

Genthon (2001a) reforça esta ideia quando ao referir-se à transferência de aprendizagem, sublinha a conveniência dos alunos trabalharem com ferramentas metacognitivas, de gestão, controlo e de avaliação das suas próprias aprendizagens.

Por outro lado, Fernandes (2005) considera que “a participação dos alunos na avaliação pode ser um processo de grande valor educativo e formativo, pois contribui para que se desenvolvam um importante conjunto de aprendizagens de natureza cognitiva e metacognitiva e de natureza social, cultural e afectiva” (p. 87).

De facto, a auto-avaliação pode ter grande valor educativo para os alunos, representar aprendizagem relevante e constituir-se como objecto de avaliação. Estes factores podem colocá-la no centro das prioridades formativas.

Alinhada com esta ideia, Roldão (2005) considera que a avaliação se pode tornar um factor de aprendizagem quando, na tarefa de avaliação, o aluno utiliza estratégias metacognitivas como a auto-análise do seu percurso. Entre outros autores, Nazario (2005), defende a tese de que a motivação e o esforço aumentam quando os professores dão aos alunos a oportunidade de avaliar os colegas e de se autoavaliar. No entanto, Pacheco (1994) destaca que, dada a natureza complexa da auto-avaliação, são necessários pré-requisitos para esta abordagem da avaliação formadora.

Genthon (2001b) defende que a auto-avaliação pode ter estatuto de objectivo de aprendizagem, mas também estatuto de objecto de avaliação. Para o autor, quando os alunos fazem auto-avaliações e analisam o seu trabalho, isso permite manipular os critérios de avaliação. Também chama a atenção para o facto de não se dever reduzir a auto-avaliação à atribuição de uma nota. De facto, a auto-avaliação representa a possibilidade de dar sentido ao que é vivido na aprendizagem (Vial, 2001a) e é ela que “transforma o aluno no actor principal da aprendizagem” (Pacheco, 1994, p. 111).

Segundo a OECD (2008), os alunos que mais frequentemente têm melhores resultados no PISA são aqueles que utilizam estratégias metacognitivas. A importância que as competências socioafectivas e metacognitivas desempenham na aprendizagem, a forma como se gerem e utilizam conhecimentos e não tanto os conhecimentos em si, pode diferenciar os alunos mais fracos dos mais fortes na resolução de problemas (Fernandes, 2005). Para McTighe e O’Connor (2005) e Harlen e Gardner (2010), os alunos que mais eficazmente aprendem são os que estabelecem metas pessoais de aprendizagem, utilizam estratégias já comprovadas, e participam na autoavaliação e na avaliação entre pares,

dominando e compreendendo o critério de qualidade que vai ser aplicado no seu trabalho. Sublinham, também, que os professores devem ajudar a cultivar estes hábitos.

Mas, para que os alunos possam apreciar o seu trabalho e orientá-lo de forma a terem melhores desempenhos os professores devem elaborar critérios de avaliação e torná-los claros e explícitos para os alunos (Fernandes, 2005; McTighe & O'Connor, 2005; Gerard, 2006). Infelizmente, em Portugal, de acordo com a análise de Fernandes (2005) a um conjunto de investigações, poucos professores, de forma isolada e sem reflectir uma acção concertada nos órgãos da escola, explicitam junto dos alunos os critérios de avaliação. Muitas vezes até são confundidos com critérios de classificação.

Nesta perspectiva, parece necessária uma atitude diferente dos professores em matéria de avaliação, nomeadamente, partilhando e negociando com outros professores, alunos e, eventualmente, encarregados de educação. Para Pacheco (1994) e Fernandes (2005), as tarefas de avaliação devem ser partilhadas com outros professores, assistentes sociais e psicólogos e o processo deve ser negociado com alunos e encarregados de educação. No entanto, para que haja uma avaliação negociada e construída tem de haver disponibilidade da parte dos professores para partilhar o poder que a avaliação lhes confere (Fernandes, 2005).

#### C) Uma Avaliação a valorizar os aspectos de natureza sócio-afectiva.

De acordo com V. De Landsheere (1988), deixar para segundo plano as componentes afectivas da competência produzirá adultos incompetentes. Para a autora, o sucesso nos estudos e na vida adulta não está só dependente das competências cognitivas ou psicomotoras, do potencial que elas representam, mas sobretudo da vontade de mobilizar outros recursos que envolvem qualidades afectivas como, paciência, perseverança, coragem, imaginação, empatia e flexibilidade.

“Affect describes a medley of noncognitive variables such as a person's attitudes, interests, and values” (Popham, 2003, p. 106). O autor prefere, mesmo, os objectivos neste domínio aos cognitivos, referindo que aluno que gosta de ler, continuará sempre a ler e quem aprendeu a gostar de discursar tem mais hipóteses de o vir a fazer melhor na vida real e sublinha que a principal razão para que os professores estejam interessados em avaliar estas variáveis é porque representam um excelente prognóstico do comportamento futuro dos alunos.

Além disso, a natureza social das aprendizagens e o seu valor ao serviço duma cidadania empenhada devem ser tomados em conta. Como refere Sternberg (2008), os cidadãos activos e empenhados devem ser capazes de executar as suas ideias e persuadir os outros do seu valor e, mais importante de tudo, ser capazes de usar o seu conhecimento sabiamente de forma a evitar más lideranças. O autor refere que não faltam líderes

egocêntricos, irrealistas, pensando que sabem tudo e eticamente pouco empenhados, embora sejam inteligentes e com conhecimentos, ou tenham sido formados em escolas prestigiadas.

Para Fernandes (2005), “as aprendizagens são processos eminentemente sociais. Ou seja, as aprendizagens, apesar de exigirem trabalho individual de interiorização, não podem ser compreendidas sem ter em conta o seu contexto social e o seu conteúdo social” (p. 26).

Neste quadro, concluímos que uma avaliação que valorize estes aspectos deve implicar a mobilização de recursos que impliquem qualidades afectivas, um comportamento moral e cívico e a necessidade de ligação aos outros.

#### D) Uma Avaliação a valorizar o trabalho colaborativo

As competências sociais como a comunicação e colaboração também fazem parte do conceito mais geral de competência (Gulikers et al., 2004). De facto, para se conseguir ter sucesso na vida é necessário saber trabalhar em equipa. Mas um dos problemas com a avaliação, que mais frequentemente é usada, é que ela não avalia o trabalho em equipa, pelo menos não directamente (Sternberg, 2008).

Fazendo o ponto da discussão teórica, Herrington e Herrington (2006) sublinham que a colaboração é uma característica da autenticidade e foi definida como o compromisso mútuo, num esforço coordenado para resolver em conjunto um problema ou criar um produto que não pode ser realizado independentemente.

Numa outra perspectiva, Gulikers et al. (2004) discordam que a colaboração seja uma característica da autenticidade, vêem-na, todavia, como uma exigência do mundo actual e referem que trabalhar em conjunto é frequentemente mais a regra que a excepção. No entanto, os autores argumentam que se a situação real pedir colaboração a avaliação deve contemplar isso, mas se, pelo contrário, solicitar desempenho individual a avaliação deve ser individual.

No entanto, a questão que se coloca é a de saber como pode a avaliação melhor reflectir as competências que interessam, não só na escola, mas ao longo da vida (Sternberg, 2008). De acordo com King e Behnke (2005), é previsível que os alunos ao longo da vida tenham que trabalhar em grupo com muita frequência, o que torna a aprendizagem em grupo ecologicamente válida. Para os autores, a educação deve preparar para a vida profissional e aqueles que não adquiram competências necessárias para trabalhar em equipa, ou que não valorizem o papel de um membro ou de um líder de uma equipa, estão em desvantagem numa sociedade que dê importância ao trabalho cooperativo.

Crahay (2002) defende que a solidariedade mútua e o respeito pelas diferenças devem substituir a perspectiva individualista e competitiva e sublinha que é preciso tolerar as desigualdades que favorecem os desfavorecidos, num mundo onde tudo é desigualdade. Para o autor, os alunos que aprendem mais rápido podem, inclusivamente, ser utilizados como tutores dos mais lentos, numa dinâmica que não prejudica os melhores e pode trazer benefícios a ambos.

Na verdade, o grupo só tem sucesso com a contribuição dos seus membros, resultante da percepção de que todos estão ligados numa interdependência positiva (Hicks, 2007), sendo que “nas estruturas sociais de objectivo cooperativo, as metas pessoais estão de tal forma ligadas que um indivíduo só pode alcançar o seu objectivo se os outros também o conseguirem” (Crahay, 2002, p. 400).

Aprendizagem cooperativa é o termo que abarca outro tipo de designações, conhecidas como aprendizagem colaborativa, aprendizagem de grupo e aprendizagem entre pares (Hicks, 2007). Segundo Hansen e Stephens (2000), neste tipo de abordagem o aluno tem o papel de principal agente de aprendizagem, toma mais iniciativa e faz uma aprendizagem socialmente interactiva na relação com outros alunos.

No entanto, aprender a trabalhar em equipa exige aprendizagem. A avaliação deve não só reflectir essa competência, como também, estar ao serviço dessa aprendizagem. Como destaca Sternberg (2008), para trabalhar em equipa e desenvolver este tipo de competência é necessário aprender e refere, como exemplo, o que fazem os membros de uma orquestra ou de um grupo de dança.

Dar por adquirido que os objectivos positivos estão associados às realizações de grupo pode representar um grande equívoco (King & Behnke, 2005). Fazendo o ponto da discussão teórica sobre o trabalho colaborativo na Educação Física, Dyson, Griffin e Hastie (2004) referem que se engana quem pensa que, simplesmente, basta colocar os alunos em grupos e pedir-lhes que colaborem que eles vão aprender ou estabelecer relações positivas entre si. Quando falha a comunicação e as competências de trabalho em equipa tudo falha (King & Behnke, 2005). Ora, as competências sociais necessárias à actividade em grupo não estão igualmente desenvolvidas em todos os alunos e não se espere que floresçam só porque os alunos fazem parte de um grupo de aprendizagem cooperativa (Hicks, 2007).

Na sua leitura da matéria publicada, King e Behnke (2005) defendem que, por um lado, é um problema nem sempre existir um currículo específico que ajude os alunos a desenvolverem estas competências e, por outro lado, muitas vezes falta tempo ou conhecimento ao professor para os preparar. Para os autores, para trabalhar em equipa com eficácia é necessário treino de como trabalhar em equipa. Todavia, como explica

Bracken (2002), a avaliação só funcionará se os alunos aprenderem as competências necessárias para terem sucesso como membros de uma equipa<sup>10</sup>.

Por outro lado, classificar o desempenho de um grupo é uma tarefa muito complicada. Existem problemas que lhe estão associados que são de difícil solução. Fazendo a síntese dos problemas referidos por Pitt, King e Behnke (2005) sistematizam-nos da seguinte forma:

- Qualquer método de seleccionar grupos e atribuir projectos dá a alguns alunos, uma vantagem e, a outros, uma desvantagem. Se todos os elementos do mesmo grupo tiverem a mesma classificação, os mais fracos podem contribuir menos.

- A contribuição para o grupo pode ser difícil de definir e avaliar. Um aluno que contribua menos pode fazer uma apresentação confiante e receber melhor nota que outro aluno com melhor contributo para o processo ou resultado

Segundo Brooks e Ammons (2003), existem problemas quando um ou mais membros de um grupo não partilham justamente o trabalho no projecto em que estão envolvidos, sendo que é frequente, o professor não conseguir observar em primeira-mão as contribuições de cada membro do grupo para o projecto. Para os autores, a maneira mais eficaz de reduzir este fenómeno é realizar avaliações desde o início e em vários pontos durante os projectos e introduzir, desde cedo, a avaliação entre pares, o que proporciona mais prática aos alunos, podendo, até, aumentar a fiabilidade dessas avaliações. Também King e Behnke (2005), consideram uma abordagem interessante os alunos atribuírem uma parte da classificação.

Finalizando este ponto, podemos seguir com Brooks e Ammons (2003) que propõem um instrumento de avaliação de grupo caracterizado por três aspectos: implementação antecipada, avaliação em vários pontos e uso de critérios de avaliação específicos.

#### E) Uma Avaliação integrada no processo de ensino-aprendizagem

A missão do currículo mantém-se constante e consiste em mostrar as habilidades, conhecimentos e atitudes que queremos que os alunos aprendam (Popham, 2003), que são tidos como necessárias numa época e contexto (Roldão, 2005), e que traduzem prioridades

---

<sup>10</sup> Na perspectiva de Bracken (2002) as principais competências do trabalho em equipa são:

1 – Delegar responsabilidades e aceitar a orientação dos seus pares;

2 – Avaliar as próprias competências, permitindo que o grupo possa usar as forças e evitar as fraquezas;

3 – Comunicar eficazmente, sendo capaz de ouvir e analisar outras opiniões e articular os seus pensamentos;

4 – Organizar o trabalho em grupo para conseguir cumprir prazos;

5 – Aceitar a crítica dos pares e conseguir melhorar o seu desempenho baseado nessas críticas e avaliar o trabalho dos outros fazendo crítica construtiva.

6 – Pôr os objectivos do grupo acima dos objectivos pessoais, já que não há sucesso individual sem sucesso da equipa.

sociais (Goodson, 1997). Ora, o momento crucial da relação entre o currículo e a avaliação acontece quando os professores pretendem fazer inferências sobre a natureza das habilidades, conhecimentos e atitudes dos seus alunos (Popham, 2003).

Zabalza (1987) sublinha que a avaliação não está separada do processo de ensino-aprendizagem, deve integrar-se nesse processo, e está ordenada, articuladamente, num conjunto de fases. Para Fernandes (2005) o ensino, a aprendizagem e a avaliação devem formar um ciclo articulado e coerente, sendo que as tarefas prescritas aos alunos devem integrar as estratégias de ensino mais adequadas, serem meios de aprendizagem e terem associado um processo de avaliação.

McTighe e O'Connor (2005) consideram que se devem apresentar tarefas de avaliação aos alunos antes de uma nova unidade ou curso. Segundo os autores, tal prática clarifica os alvos tanto para os professores como para os alunos, sendo que a apresentação das tarefas de avaliação não só fornece um objectivo de aprendizagem com significado, como permite que os alunos vejam uma razão para a sua aprendizagem. Gulikers et al. (2004) são autores, entre outros, representativos da ideia que a tarefa de aprendizagem deve ser semelhante à tarefa de avaliação, embora os objectivos sejam diferentes. Além disso, defendem que os alunos precisam de ter a oportunidade de praticar as situações de prova, antes de serem usadas como avaliação.

Por outro lado, deve-se tomar em conta que não se consegue avaliar todos os domínios do currículo ou o essencial de cada um, nem avaliar tudo o que um aluno sabe e é capaz de fazer (Fernandes, 2005). O conselho de Popham (2003) é que se estabeleçam prioridades, preferencialmente, em equipas colegiais de professores. Para este autor, a avaliação adequada dos conteúdos considerados altamente prioritários conduzirá a melhores decisões de instrução porque há um aumento de clarificação proporcionada por esta operacionalização da avaliação, quer quanto à natureza, quer quanto aos propósitos do currículo.

Em síntese, podemos seguir com a afirmação de Gulikers et al. (2004): “learning and assessment are two sides of the same coin, and that they strongly influence each other” (p. 68). De acordo com estes autores, é necessário que a instrução autêntica esteja alinhada com avaliação autêntica para a aprendizagem mudar de rumo em direcção ao desenvolvimento de competências. Na mesma linha, também Sternberg (2008), concorda que a instrução e a avaliação são os dois lados da mesma moeda, não são duas moedas e destaca o papel que a avaliação deve ter na condução do processo de instrução.

### **1.3.2. A Avaliação Autêntica**

As chamadas avaliações alternativas, onde se integra a avaliação autêntica, contêm as características enunciadas anteriormente. Partindo dos traços gerais



identificadores das avaliações alternativas analisemos com mais detalhe as características próprias da avaliação autêntica.

De acordo com Fernandes (2006, 2008), a expressão avaliação alternativa é uma espécie de guarda-chuva que a literatura utiliza para abrigar os processos de avaliação que se destinam a regular e melhorar as aprendizagens. Segundo o autor, este tipo de avaliação orienta-se mais para ajudar os alunos a aprender do que para classificá-los, utiliza situações em contexto e integradas no processo de ensino aprendizagem que não se baseiam nos tradicionais testes de papel e lápis, e em que os alunos assumem papéis de maior importância em todo o processo. Tem características e práticas que se associam à avaliação formativa que se inspira nos princípios que decorrem do cognitivismo e do construtivismo (Fernandes, 2005, 2006).

Segundo a perspectiva que valoriza a autenticidade das tarefas e contextos de avaliação, Gulikers et al. (2004) concluem na sua análise da literatura, destacando que as principais características da avaliação alternativa se traduzem no facto dos alunos terem responsabilidade na sua própria aprendizagem e, além disso, a avaliação implicar contextos e tarefas que tenham como referencia a vida real. Para os autores, os momentos de avaliação e métodos utilizados devem ser múltiplos para se conseguir avaliar a aprendizagem e o desenvolvimento do aluno. Nazario (2005), por seu turno, considera que algumas das avaliações alternativas também têm como característica oferecerem aos alunos a oportunidade de dar mais que uma resposta ou de resolver o problema de mais de uma maneira.

Uma das formas alternativas de avaliação é a chamada «avaliação baseada no desempenho» que está em sintonia com a ideia de que uma competência “não se pode observar directamente a não ser através de uma actuação ou de um produto observável, ou seja, de um desempenho” (Serpa, 2010, p.105).

Meisels et al. (1995) explicam que este tipo de avaliação, surgindo como resposta daqueles que criticavam a avaliação descontextualizada, fornece um método para avaliar os elementos de aprendizagem e desenvolvimento que os testes tradicionais têm dificuldade em integrar, como as habilidades de pensamento de ordem superior, onde se incluem a análise, a síntese, a avaliação e a interpretação de factos e ideias, mas também, a iniciativa e a criatividade. Para os autores, esta avaliação envolve os alunos em tarefas metacognitivas de revisão e avaliação da sua própria aprendizagem, fazendo a ponte entre várias áreas do currículo e “permitem aos alunos demonstrar os seus conhecimentos ou habilidades para resolver problemas, fazer computações matemáticas, escrever jornais ou ensaios, conduzir experiências, apresentar relatórios orais, ou elaborar um portfólio representativo do trabalho efectuado” (p. 278). Para Popham (2003), este tipo de avaliação

implica sempre a realização de uma tarefa e só é recomendada para os objectivos curriculares prioritários, sendo que, por vezes, até estão em causa competências muito sofisticadas.

Já anteriormente, Eisner (1999) tinha assinalado que, na avaliação baseada no desempenho, os alunos dão provas daquilo que sabem e conseguem fazer em situações reais, sendo que só através do seu desempenho nessas situações se podem fazer inferências válidas.

Neste quadro, em que há autores que referem o realismo das situações de avaliação como uma característica deste tipo de avaliação, será que as designações «avaliação baseada no desempenho» e «avaliação autêntica» querem dizer o mesmo e são intermutáveis?

Moon et al. (2005) sublinham que é importante fazer a distinção entre avaliação baseada no desempenho e avaliação autêntica porque vários autores as vêem como sinónimos. Gulikers et al. (2004), fazendo o ponto da discussão teórica, afirmam que na avaliação baseada no desempenho a fidelidade, da tarefa e das condições em que o desempenho deve normalmente ocorrer, não é uma questão muito importante e já na avaliação autêntica é absolutamente determinante representando a grande diferença entre as duas. Assim, concluem que qualquer avaliação autêntica é uma avaliação que está baseada no desempenho, mas nem todas as avaliações baseadas no desempenho são autênticas.

Na perspectiva inicial de Wiggins (1990), a avaliação autêntica deve solicitar desempenhos eficazes em tarefas que espelhem as grandes prioridades das actividades de instrução como pesquisar, escrever, analisar artigos ou colaborar com outros em debates, assumindo desafios e papéis que ajudem os alunos a estar preparados para enfrentar as ambiguidades complexas da vida adulta e profissional. Para este autor a validade de uma prova de avaliação autêntica depende da forma como simulando o mundo real se solicitam as competências que são necessárias usar na realidade.

Para Gulikers et al. (2004), é a situação critério na vida real que baliza a autenticidade. Quanto maior for a semelhança entre as competências ou combinações de conhecimentos, habilidades e atitudes requeridas pela avaliação e as que são necessárias na situação critério, maior nível de autenticidade se consegue, sendo este o campo da avaliação autêntica. Uma avaliação que requer realismo da tarefa e do contexto.

De acordo com Stobart (2010), manter a avaliação tão autêntica quanto possível significa avaliar directamente as competências, estar perto da validade de constructo. Quanto mais autêntica for a avaliação mais confiança podemos ter que uma classificação representa as competências da vida real e a compreensão que se supõe ela avaliar.

Gulikers et al. (2004) consideram a autenticidade um elemento importante nas novas formas de avaliação e defendem que aumenta a validade preditiva devido à semelhança entre a avaliação e a prática profissional real. Por outro lado, na sua leitura da matéria publicada, referem que o resultado da avaliação deve ser um produto ou desempenho de qualidade que possa ser pedido na vida real e deve ser apresentado para outros, oralmente ou por escrito, para garantir que o seu domínio é genuíno.

Herrington e Herrington (2006) embora concordando que todos os cenários devem ser baseados em problemas e que esses problemas devem ser reais, argumentam que não bastam contextos com problemas reais para garantir a autenticidade, é necessário um realismo cognitivo da tarefa.

Gulikers et al. (2004), fazendo o ponto da discussão teórica, consideram que a autenticidade da avaliação se encontra na similaridade entre as demandas cognitivas da avaliação e as demandas cognitivas na situação critério em que a avaliação é baseada, mas argumentam que não basta lidar somente com o pensamento requerido, sendo necessário desenvolver competências porque a vida real pede que se integre e coordene conhecimento, habilidades e atitudes e que se seja capaz de aplicá-los em novas situações.

Numa perspectiva que integra as várias componentes da avaliação, Herrington e Herrington (2006), revendo a literatura, referem que a autenticidade manifesta-se no fluxo de si própria, e não é uma característica objectiva de qualquer componente isolada como aluno, tarefa ou envolvimento. Concluem que a autenticidade está nas interacções dinâmicas entre as várias componentes.

Por outro lado, discutindo a forma como os alunos percebem a avaliação e a possível influência na sua aprendizagem, Gulikers et al. (2004) consideram ser a autenticidade subjectiva e aquilo que os alunos percebem como autêntico não ser necessariamente o mesmo que os professores vêem como autêntico, o que faz as percepções importantes para que a avaliação influencie a aprendizagem. Para os autores, a percepção de autenticidade também muda em consequência do nível de educação, interesses, idade ou experiência e defendem que não se deve perder de vista o nível educacional dos alunos, já que os mais fracos podem não ser capazes de lidar com a autenticidade numa situação real, que seja complexa.

Em síntese e procurando clarificar conceptualmente a avaliação autêntica entendemo-la como a avaliação que usa as mesmas competências, ou combinações de conhecimentos, habilidades e atitudes, que são necessárias na actividade cultural referente.

De acordo com Serpa (2010), o referente corresponde ao ideal que se têm em relação aos objectos a avaliar. É um elemento exterior que orienta o sentido das apreciações a realizar.

Nesta perspectiva, é a actividade cultural referente que estabelece as características do objecto de aprendizagem e, portanto, de avaliação, na sua forma e conteúdo e, também as fronteiras da autenticidade. Esta é uma orientação para o domínio do objecto culturalmente significativo, que se apresenta no currículo como matéria e objectivo e que, assim, determina se as tarefas e o contexto da avaliação são «reais».

Trata-se, assim, de um tipo de avaliação que está baseada no desempenho em situações problemáticas representativas do constructo, onde a mobilização de recursos integrados necessários para responder às situações problema exige muito mais do que a exclusiva participação do domínio cognitivo.

Mas este esforço de qualificação da avaliação representado pela avaliação autêntica debate-se com os problemas característicos das avaliações alternativas.

Harlen e Gardner (2010) consideram que para a avaliação ser válida é necessário uma alternativa aos testes escritos em que os alunos sejam confrontados com problemas reais e obrigados a ligarem-se a outros.

Moon et al. (2005) defendem que, para o aluno demonstrar as competências desejadas, a avaliação deve fornecer múltiplas evidências, permitindo interpretações multidimensionais. Para os autores, a avaliação deve estar focada na essência da tarefa e não no que é fácil de pontuar.

Só que garantir estes pressupostos não é tarefa fácil. Nas avaliações alternativas que se consideram autênticas e estão baseadas no desempenho, os problemas mais difíceis de resolver estão relacionados com os custos da avaliação, particularmente o tempo despendido por professores e alunos, e com a pouca fiabilidade de muitas classificações. De acordo com Fernandes (2005), os autores que defendem a avaliação alternativa recomendam as questões abertas tipo problemas, tarefas ou ensaios. A dificuldade está em as provas alternativas serem mais difíceis de elaborar, demorarem mais tempo a elaborar, administrar e classificar que as outras (Nazario, 2005; Fernandes, 2005) e apresentarem, por vezes, problemas ao nível da fiabilidade das classificações (Nazario, 2005).

De facto, ao incluir competências de pensamento de nível superior para aumentar a validade, a fiabilidade pode ficar mais perto de falhar porque estes aspectos não são facilmente avaliados (Harlen, 2004).

Em situações de aprendizagem mais abertas é difícil especificar os objectivos, estando neste caso as competências de ordem superior, que são mais difíceis de operacionalizar em termos de comportamentos observáveis (Allal, 2001). Nas aprendizagens complexas, para atingir um objectivo conjugam-se factores cognitivos e afectivos que são difíceis de fechar num qualquer item de um teste (V. De Landsheere, 1988).

Nesta mesma linha, De Ketele e Gerard (2005) explicam que as provas baseadas em competências contêm situações complexas que exigem uma produção complexa para a qual é necessário um certo tempo de resolução. Inclusivamente, a produção pode ter que ser decomposta em várias tarefas, compreendendo elas mesmas várias etapas, com várias formas de resolução possíveis. Não é possível constituírem-se itens, porque isso corresponderia a uma nova decomposição da tarefa complexa e das suas unidades, o que não permitiria mais avaliar a competência que consiste, precisamente, em gerir bem a complexidade.

Na análise do problema, Popham (2003) sublinha que é necessário fazer muitas provas baseadas no desempenho antes que se possam fazer inferências válidas sobre o domínio das habilidades discretas.

Segundo Gerard (2006) colocam-se, então, questões complicadas para resolver. Como validar as provas? Como garantir a equivalência de provas que avaliam a mesma competência? Como as corrigir?

De acordo com De Ketele e Gerard (2005), as técnicas de validação destas provas ainda estão por inventar, com outros pressupostos diferentes dos da teoria clássica. Não se podem validar os instrumentos como se fazem nas provas clássicas que têm os conteúdos como referência e estão baseadas numa abordagem psicométrica<sup>11</sup>. Numa prova de competências, há uma produção complexa que é analisada e apreciada por um júri ou por um avaliador, segundo critérios claros.

Para estes autores, os itens são situações problema concretas e o seu universo de referência é a família de situações a que corresponde a competência. Como a questão da equivalência de situações, que são por definição complexas, é muito difícil, senão impossível, de resolver, preconizam, então, que se utilizem juízes para verificar se as situações correspondem à família a que pertencem. Para se fazer bem este trabalho é indispensável dispor de uma descrição suficientemente precisa da família de situações e de uma definição de parâmetros que a caracterizam.

Por outro lado, vários autores parecem concordar que o uso de rubricas configura um bom método para ajudar os professores a avaliar o desempenho dos alunos.

Segundo Sternberg (2008), a melhor forma de avaliar respostas a questões que parecem subjectivas é com rubricas bem desenvolvidas. McTighe e O'Connor (2005) referem que as rubricas bem elaboradas orientam os professores na avaliação do trabalho

---

<sup>11</sup> Na lógica da psicomетria, usando uma avaliação de referência normativa, espera-se que os resultados dos alunos se distribuam normalmente numa curva como a de Gauss, em que a média se confunde com a moda e a mediana. O índice frequentemente utilizado é o índice de discriminação que visa verificar que os itens da prova discriminam os alunos, os fortes dos fracos. Na lógica da distribuição normal 27% estará na categoria superior e 27% na categoria inferior (De Ketele & Gerard, 2005).

dos alunos, já que traduzem o mais importante, ou seja, as qualidades que um produto ou desempenho devem ter.

Para Popham (2003), uma rubrica é um guia de classificação, sendo que o critério de avaliação é o seu componente mais importante e deve ser acompanhado por qualidades que o definem. Para o autor são as rubricas que determinam se a cotação é mais alta ou mais baixa e que reduzem a possibilidade dos critérios poderem ser mal percebidos. Moon et al. (2005) consideram que, no que diz respeito às rubricas, tudo começa com a revisão dos objectivos a avaliar e a identificação dos elementos de desempenho a ser avaliados.

De acordo com Popham (2003), a estratégia de classificação pode ser holística ou analítica. Na holística a classificação deve tomar em consideração quão bem a resposta satisfaz todo o critério de avaliação. É uma abordagem rápida e pouco dispendiosa e por isso a preferida em operações de classificação em larga escala. Na analítica a classificação é apurada num resultado final que juntou todas as cotações em função de apreciações critério a critério.

Segundo McTighe e O'Connor (2005), quando se usam rubricas comuns, os resultados da avaliação são mais consistentes porque os critérios de desempenho não variam consoante o professor, o departamento ou a escola e, para além disso, podem ajudar os alunos a serem mais eficazes e honestos na autoavaliação. Para os autores, as divergências entre professores e alunos podem representar a oportunidade de discutir os critérios de avaliação e assim, virem a alinhar as suas apreciações. Para Popham (2003), alunos que dominem os critérios de avaliação podem avaliar os seus desempenhos e os dos colegas e isto ser uma componente da instrução.

As avaliações alternativas, em particular, as avaliações autênticas que são baseadas no desempenho colocam problemas novos que são difíceis de resolver. Nada disto é simples ou fácil e para complicar tudo ainda não há muita experiência acumulada que suporte a complexidade de tais tarefas. Estamos, no entanto, convictos que das características do professor do futuro farão parte o talento, a intuição e o conhecimento para fazer bem feito aquilo que agora se preconiza. Na perspectiva inicial de Wiggins os ganhos da avaliação autêntica superam os seus custos e traduzem-se em desenvolvimento profissional dos professores, aprendizagens dos alunos e mais e melhor avaliação local. Como refere Sternberg (2008), esta avaliação custa mais tempo e dinheiro. Mas se considerarmos os benefícios o seu custo é menor.

#### *A Avaliação Autêntica numa nova Cultura de Escola*

Analizadas as diferentes características que podem estar associadas às avaliações alternativas e analisadas as especificidades, próprias, da avaliação autêntica, há um aspecto que se assume como determinante para garantir coerência entre os professores, o

trabalho colaborativo entre eles. Se não vejamos, avaliar o mesmo desempenho de forma diferente é muito provável que aconteça quando diferentes avaliadores estão entregues a si próprios (V. De Landsheere, 1988). Se quem avalia influencia os resultados, não se garante a consistência da avaliação (Harlen, 2007), que em Portugal é reconhecidamente débil, particularmente, na avaliação interna (Fernandes, 2005).

Ora, nas provas alternativas que são por natureza mais complexas, abertas e difíceis de classificar (Nazario, 2005; Fernandes, 2005), manifesta-se frequentemente o desacordo entre os correctores (Kellaghan & Madaus, 2003). Na verdade, “a partir do momento em que se conjuga uma série de aspectos a avaliar, a possibilidade de divergência aumenta” (Serpa, 2010, p. 106).

No entanto, ter um currículo clarificado, com provas de avaliação que reflecta o essencial desse currículo, definir critérios e instrumentos comuns de avaliação, aferir as avaliações tentando reduzir as divergências classificativas, são meios para garantir a redução da subjectividade e contributos para a emergência de uma cultura de acção colectiva na educação e, em particular, na avaliação. Para Harlen (2007), os professores só aumentarão a fiabilidade das suas avaliações e alinharão as suas apreciações sobre o desempenho dos alunos, se trabalharem em conjunto.

Só que a solidão em matéria de avaliação (Clímaco, 2005; Fernandes, 2005; Fernandes et al., 1996), como domina os hábitos e comportamentos dos professores, implica, necessariamente, mudanças na cultura de cada escola. De acordo com Fernandes (1994, 2005) e também, Fernandes et al. (1996), a ausência de uma cultura de avaliação é um problema que deveria ser cuidadosamente enfrentado pelas escolas e pelos grupos disciplinares.

De facto, recolher informação credível sobre as aprendizagens dos alunos e o trabalho das escolas, só se consegue com trabalho colectivo em equipas pedagógicas, e a inexistência de um estilo colectivo, impede, por um lado, que se façam apreciações rigorosas e se tomem decisões de aperfeiçoamento das escolas, dos projectos ou das pessoas, e por outro lado, torna mais permeáveis as próprias pessoas, alunos ou professores, às idiossincrasias dos avaliadores, e as escolas às avaliações descontextualizadas que originam os tão discutidos rankings.

Sintetizando o pensamento de outros autores, Gulikers et al. (2004) explicam que as duas razões mais importantes para usar avaliações autênticas baseadas em competências são a sua validade de constructo e o seu impacto na aprendizagem dos alunos, sendo este impacto referido de validade consequential.

Portanto, as provas de avaliação autêntica têm de respeitar a validade de constructo, ou seja, a essência da matéria a ser avaliada, e não o contrário que é a

adaptação da validade às provas, forçando a fiabilidade e com isso minando a validade. O constructo diz respeito, não só à matéria a ser avaliada, mas também à sua própria construção que é uma questão social na escola e deve ser resolvida em grupo. Se a autenticidade garante inferências válidas, o esforço deve ser investido na consistência da avaliação e na redução dos seus custos.

Como o trabalho colectivo e colegial é quase inexistente e os órgãos da escola não se coordenam de forma a reflectir uma acção global coerente, existe um sentimento geral que, de acordo com Fernandes (2005), se caracteriza por “relativa falta de confiança que os professores manifestam relativamente às suas avaliações e às decisões que delas decorrem, assim como aos procedimentos que utilizam” (p. 90).

Tentar alterar este estado de coisas não é fácil. Para Clímaco (2005), a cultura de escola dificilmente muda, havendo valores que não se discutem e com as mentalidades estabelecidas a terem, por vezes, mais força que as propostas de inovação. Como a avaliação não se muda por simples decreto (Pacheco, 1994) o caminho está na mudança da «cultura de escola», na capacidade de cada escola construir a sua própria cultura, procurando reduzir os constrangimentos externos (Barroso, 2005).

De acordo com Clímaco (2005), por vezes adoptam-se comportamentos defensivos face ao desconhecido e à complexidade, ou para manter a organização existente, principalmente, em momentos de grande incerteza. Para a autora, a incerteza que caracteriza a complexidade só pode ser dominada por uma visão de futuro baseada na capacidade de antecipação e de decisão.

Mudar a escola é mudar a cultura e a organização, implementando uma cultura de colaboração entre os professores (Pacheco, 1994). Como destaca Roldão (1998, 2006), são necessárias novas culturas educativas assentes no trabalho colaborativo sistemático, articulando o esforço de todos para dar resposta a situações problemáticas, aprofundando processos e resultados, verificando os pontos fortes e fracos e colocando em comum, tudo o que cada um sabe e que é específico. Também Clímaco (2005) considera importante a cooperação docente e o trabalho de equipa, aproveitando-se o talento e a inteligência de cada um para a elevação das tarefas comuns. A autora sublinha que a aprendizagem em equipa permite criar sinergias através da partilha de experiências e da capacidade reflexiva dos seus membros, num processo que tem de ser interactivo e colegial.

Por outro lado, o treino da avaliação e os processos de aferição de critérios exigem tempo de trabalho colectivo entre os professores. Para isso, é necessário repensar o tempo de escola. É necessário que as estruturas pedagógicas e de gestão da escola vejam este trabalho colectivo como importante e lhe atribuam tempo na distribuição de serviço docente, particularmente, na componente não lectiva. Sem tempo, dificilmente a avaliação pode ser



estruturante e estar ao serviço do desenvolvimento e aperfeiçoamento das escolas. Como sublinha Darling-Hammond (2004), os professores precisam de tempo para trabalhar e aprender em conjunto. São disto exemplo, o tempo para partilha, para trabalhar na net com outros colegas, para avaliar o ensino e a aprendizagem, para contactar com especialistas técnicos e recursos. Para a autora, atribuir recursos que apoiem, sistematicamente, esse trabalho e a aprendizagem, inclui a necessidade de reestruturar o tempo de escola para permitir tal oportunidade.

Assim, quer estejam a avaliar individualmente as aprendizagens dos alunos nas suas turmas, que estejam colectivamente envolvidos em provas de avaliação ao nível do grupo/ departamento, ou do sistema educativo, os professores precisam de tempo e de uma nova cultura de trabalho em equipa para garantir a validade e fiabilidade das avaliações e das inferências sobre os resultados.

Mudar a cultura de escola é, também, ultrapassar a tradicional ausência de cultura de avaliação, passando do tradicional individualismo docente em matéria de avaliação para práticas colectivas de avaliação, centradas na aferição de critérios, na interpretação da informação recolhida e na tomada de decisões que permitam agir com base nos resultados. A energia gasta nos processos de moderação, nos esforços de validação, tem benefícios que superam os custos. São mesmo muito pequenos os custos comparados com os benefícios em formação de professores, em animação pedagógica das escolas, em resultados válidos e fiáveis, e portanto, credíveis para alunos, pais, professores e decisores.

#### **1.4. Da Avaliação Tradicional à Avaliação Autêntica em Educação Física**

As práticas dos professores que se têm apoiado em correntes pedagógicas tradicionais influenciam fortemente a avaliação em EF. Também nesta área disciplinar manifestam-se, de forma expressiva, alguns dos problemas tradicionalmente associados à avaliação escolar. Wood (2003) refere que os processos de desempenho de ordem superior raramente são avaliados em condições realistas e autênticas e as técnicas de avaliação utilizadas carecem de autenticidade e de contextualização. Para o autor, as técnicas tradicionais avaliam estritamente as diferentes características dos domínios cognitivo, afectivo e psico-motor, tratando-os em separado.

“These traditional approaches to assessment in physical education reflect traditional approaches to curriculum construction and pedagogy” (Hay, 2006, p. 318).

Para este autor a pedagogia, a didáctica e os currículos tradicionais, centrados no professor e focados no desenvolvimento de habilidades discretas, aptidão motora, saúde e conhecimentos sobre a aptidão motora, têm exercido uma forte influência nas técnicas de avaliação sumativa e formal. Na sua leitura da bibliografia, as técnicas de avaliação

tradicional em EF integram os testes de avaliação da aptidão motora e os testes de avaliação da precisão da habilidade motora e ainda, outras técnicas de avaliação que incluem testes sobre regras, tática e história, e escalas psicométricas e inventários, sendo avaliações que representam principalmente avaliações de produto. Também Siedentop e Tannehill (2000) reforçam a ideia de que tradicionalmente avalia-se com testes de aptidão motora, testes de habilidades isoladas ou dando pontos subjectivamente para esforço e participação.

Alguns exemplos de testes normalizados de habilidades são os testes AAHPERD (Nazario, 2005). Justamente, Silva (2008), apoiando-se noutros autores, refere que os professores avaliam aspectos relacionados com as habilidades, inspirados em testes normalizados, como os testes AAHPERD, elaborando muitas escalas de apreciação e listas de verificação.

Por outro lado, Gréhaigne et al. (2005), revendo a bibliografia, referem que pelo facto de usarem testes similares não significa que os professores utilizem os testes padronizados que estão publicados. Combinados com testes sobre regras e regulamentos, os testes de habilidade são normalmente utilizados como forma de avaliação sumativa, repercutindo-se nas classificações dos alunos (Oslin et al., 1998).

Estas práticas de avaliação estão muito alinhadas com os métodos de ensino que os professores mais têm utilizado. Sintetizando o pensamento de vários autores, Silva (2008) destaca o facto de o ensino ter estado muito centrado na valorização das habilidades técnicas que se assumem como uma referência para o desempenho, e, por vezes, também na estética que está associada ao movimento. De acordo com Gréhaigne et al. (2005), a abordagem tradicional foca-se, em primeiro lugar, no ensino de habilidades técnicas que mais tarde são aplicadas em jogo e só depois apresenta os elementos de natureza tática. As ligações entre a técnica e o contexto de jogo são muito esporádicas (Kirk & McPhail, 2002).

Relacionando a avaliação e o ensino, Turner e Martinek (1995), argumentam que os professores apresentam as habilidades em primeiro lugar, esperando que depois sejam aplicadas na situação de jogo. Só que, frequentemente, a qualidade do jogo permanece fraca porque apesar de praticarem persistentemente as habilidades, apenas alguns alunos conseguem fazer a transferência para o jogo. Segundo os autores, alunos com excelente desempenho técnico podem não compreender quando e onde usar as suas técnicas, ou seja, não sabem o que fazer no contexto formal de aplicação desses conteúdos. Concluem que os alunos quando praticam uma habilidade querem saber quando aplicá-la, precisam de uma prática com relevância e significado.

Na mesma linha, Graça et al. (2003) afirmam que os modelos tradicionais focados na aprendizagem de elementos técnicos sem contextualização, que transformam a aprendizagem isolada num fim em si mesmo, tornam o tempo de exercitação monótono, aborrecido e sem significado. Para os autores, os melhores assumem os papéis mais relevantes e têm mais tempo de jogo enquanto que os mais fracos são frequentemente ostracizados.

Na verdade, os métodos de ensino e de avaliação, permanecem quase os mesmos (Oslin et al., 1998). Estão focados na técnica (Gréhaigine et al., 2005) e utilizam os testes que avaliam em exclusivo, a execução da habilidade. Acredita-se, erradamente, que a habilidade para jogar o jogo, utilizando adequadamente as habilidades, está representada nas pontuações desses testes (Oslin et al., 1998).

Na sua leitura da matéria publicada, Nazario (2005) explica que as avaliações se referem a um desempenho de determinado tipo, mas raramente são contextualizadas. Refere que os alunos aprendem as habilidades mas não sabem como aplicá-las no jogo. Conclui que estas habilidades ao serem avaliadas, frequentemente, duma maneira subjectiva, estão sujeitas à crítica porque há falhas na validade e na fiabilidade.

Esta autora, entre muitos outros, é representativa da ideia que a maneira escolhida por muitos professores para fazer estas avaliações leva a que alunos fiquem inactivos, razão porque talvez em muitos casos os professores não usem a avaliação formal.

Na verdade, os professores que se fiam em testes de habilidades para avaliar o desempenho do aluno sofrem dissabores. Esses testes como são descontextualizados, com situações não relacionadas com o jogo e a não considerar as dimensões sociais a ele associadas, não só não representam uma visão ampla do desempenho no jogo, como também, não vaticinam o desempenho no próprio jogo (Griffin et al., 1997; Oslin et al., 1998).

Nesta perspectiva tradicional de ensino predomina a tendência para avaliar a habilidade para executar a técnica isolada e não a habilidade para jogar o jogo. Acredita-se que uma vez dominadas as habilidades fundamentais, elas conseguem ser adequadamente aplicadas em contexto de jogo. Mas, como destaca Crespo (1979), não se pode afirmar que alguém tem muita habilidade para o futebol só porque dá toques de bola, com os pés ou a cabeça, fora do próprio jogo.

Por outro lado, tal como nos JDC, também os métodos análitos para avaliar e ensinar no conjunto da EF, parecem permanecer como tendência dominante. Como referem Dejong, Kokinakis & Kuntzleman (2002), durante muitos anos os programas pedagógicos estavam centrados na participação em actividades e não no ganho de competências ao nível da execução motora, do desenvolvimento das capacidades físicas, ou da

aprendizagem de processos que permitam desenhar um plano pessoal de condicionamento físico.

Considerando os trabalhos sobre esta questão, Nazario (2005) explica que na EF, os alunos têm sido classificados mais pela presença ou participação nas aulas do que pelas suas conquistas na aprendizagem, sendo que os métodos mais comuns de classificação revelam que os professores confiam em avaliações subjectivas centradas na presença, participação, esforço ou comportamento. Conclui que a principal prerocupação pedagógica parece ser que os alunos se mantenham em actividade e não avaliar o seu progresso nas aprendizagens.

No entanto, para Fay e Doolittle (2002), há agora a responsabilidade de demonstrar que os alunos aprenderam as matérias, defendendo que deve prestar-se contas sobre os resultados alcançados e não somente sobre a participação dos alunos. Também Nazario (2005) critica o facto de os professores usarem a avaliação com o objectivo de classificar mais do que analisar criticamente o processo de instrução ou o conteúdo do currículo. A autora assinala que os professores necessitam ver a avaliação como parte da instrução, não só para classificar, mas também para melhorar o desempenho dos alunos.

Neste quadro, em contraponto com as formas tradicionais de avaliação, tal como no conjunto da educação também na EF as formas alternativas de avaliação começaram a ganhar espaço na prática dos professores, em particular, a linha promissora da avaliação alternativa que se institui em práticas que se pretendem autênticas.

#### *Avaliação Autêntica em Educação Física*

A avaliação alternativa ganhou esta designação porque se constituiu como alternativa às formas mais tradicionais de avaliação (Wood, 2003; Siedentop & Tannehill, 2000). É um tipo de avaliação com mais significado e que reconhece o sucesso e os progressos individuais (Nazario, 2005). De entre os muitos exemplos de avaliação alternativa em EF, esta autora destaca: resumos e/ou sínteses de artigos publicados sobre actividade física; os ensaios escritos; as apresentações orais; a avaliação entre pares; as demonstrações publicas; os fóruns de discussão; os planos de condição física; os portefólios, em que se pode registar as evidências das outras avaliações alternativas; as avaliações baseadas no desempenho; e as avaliações autênticas.

Fazendo o ponto da discussão teórica, Wood (2003) afirma que as avaliações alternativas em que o aluno demonstra o comportamento desejado pelo avaliador, são as avaliações baseadas no desempenho, a «performance». De acordo com Doering (2000), a avaliação baseada no desempenho foca-se no pensamento de ordem superior e em habilidades complexas. Para o autor, este tipo de avaliação deve implicar a resolução de problemas complexos que solicitem diferentes tipos de comportamento ou performance, um

esforço significativo, utilizando soluções contextualizadas em tarefas abertas. Além disso, explica que estas avaliações podem ser feitas individualmente ou em grupo e devem envolver um significativo grau de escolha do aluno.

Por seu turno, Doering (2000) refere que a avaliação baseada no desempenho pode tomar a forma de respostas construídas, produtos, performances, ou processos, dando como exemplos de “respostas construídas” os ensaios e as respostas curtas. Como exemplos de “produtos”, indica os livrinhos, as brochuras, os posters, as exposições e os vídeos. São exemplos de “performances” as coreografias de ginástica ou dança e os jogos. Finalmente, para exemplificar os “processos”, aponta as “entrevistas” ou as descrições da elaboração de uma determinada resposta.

Temos ainda, noutra perspectiva, as avaliações autênticas que aproveitam as situações reais (Harvey, 2006) e que requerem um desempenho num contexto que é similar aquele em que as competências serão, eventualmente, usadas (Siedentop & Tannehill, 2000). Estas avaliações têm grande aceitação naqueles que têm uma perspectiva cognitivista, construtivista ou situada da aprendizagem, que reconhecem a interdependência do currículo, da pedagogia e da avaliação (Hay, 2006).

Fazendo o ponto da discussão teórica, Wood (2003) afirma que, mais especificamente, as avaliações autênticas são avaliações baseadas no desempenho, administradas em contextos de vida real. Siedentop e Tannehill (2000) consideram-nas holísticas e caracterizam-nas como as que acontecem em situações realistas, reflectindo a vida real e o que os alunos fazem fora da escola. Na sua leitura da matéria publicada, Harvey (2006; 2007) explica que este tipo de avaliação é também regular, formativa, contínua e comportamental, devendo o professor conseguir observar a habilidade que está a ser avaliada.

Gréhaigne et al. (2005) consideram que são as situações que têm relação com contextos de vida real que dão significado à actividade de aprendizagem, devendo os factores culturais e motivacionais terem um papel importante nas escolhas que se fazem. Siedentop e Tannehill (2000) consideram que é na variedade de tarefas e situações de avaliação que os alunos têm a oportunidade de demonstrar as suas competências e que, só crescem e aprendem verdadeiramente quando têm de aplicar o conteúdo em contexto. Para Martin, Kulinna e Cothran (2002), é importante utilizar tarefas autênticas que coloquem os alunos em situações de avaliação significativas e contextualizadas, porque eles precisam de ver a ligação entre o formato da avaliação e o que fazem na aula. Os autores sublinham que a falta de significado ocorre muitas vezes nos testes de habilidade formais porque são realizados fora de contexto.

Na perspectiva de Hay (2006), as tarefas contextualizadas típicas da avaliação autêntica devem ser significativas para solicitarem o conhecimento de ordem superior e para permitir a transferência das aprendizagens para contextos para além da sala de aula.

Discutindo essa orientação, Nazario (2005) explicou que dar ênfase ao pensamento de nível superior e às aprendizagens mais complexas, implica uma avaliação que se interesse mais por o aluno compreender verdadeiramente o que está a ser ensinado e não tanto em memorizar para testes escritos. A autora salienta que não é suficiente saber regras e estratégias, é também necessário saber como aplicá-las em diferentes situações, e conclui que as tarefas apresentadas têm que ter significado, ou seja, representar o desempenho no campo.

Como referem Gréhaigne et al. (2005), “the whole set of elements can come into play only at a superior cognitive level where one can make choices, face unexpected events, and find solutions for new situations” (p. 46).

Em síntese, as avaliações autênticas em EF devem estar baseadas no desempenho, serem necessariamente significativas e contextualizadas com referência à vida real e solicitarem o pensamento de nível superior e as aprendizagens complexas.

Mas também devem estar alinhadas e integradas no processo de ensino e aprendizagem e contarem com a participação activa dos alunos.

De acordo com Graça e Mesquita (2007), os desempenhos contextualizados e o alinhamento da instrução com a avaliação são marcas da avaliação autêntica. Para os autores, as práticas de avaliação autêntica acontecem durante o processo de aprendizagem e não o interrompem. Revendo a literatura, Nazario (2005) defende que deve haver uma ligação entre as tarefas diárias de instrução e a avaliação, que a avaliação está tão profundamente ligada com a instrução que por vezes não se distingue dela, ou seja, na EF genuína avaliam-se os alunos durante as aprendizagens. Para Gréhaigne et al. (2005), se a avaliação estiver verdadeiramente integrada no processo de ensino-aprendizagem deve ser realizada em contexto, não perturbando o normal funcionamento da aula.

“Teachers interested in authentic assessment need to be equipped with instruments that can be used while teaching and observing the students within the lesson” (Harvey, 2004, p. 20). Estes Instrumentos de avaliação desenvolvidos para que os professores possam avaliar e ensinar ao mesmo tempo devem ser válidos e fiáveis (Harvey, 2006). Inclusivamente, Siedentop e Tannehill (2000) defendem que a integração da avaliação no processo de ensino aprendizagem deve passar por se ensinar através da prova. Para os autores, os alunos devem praticar o que se quer que aprendam e o que vai ser avaliado, embora também devam ser utilizadas outras tarefas consideradas facilitadoras e benéficas para a aprendizagem.

Por outro lado, Gréhaigne et al. (2005) defendem a participação activa dos alunos numa avaliação que esteja integrada no processo de ensino-aprendizagem. Na sua leitura da matéria publicada, Nazario (2005) afirma que os alunos devem saber como vão ser avaliados e que essa Informação, *a priori*, dos critérios de avaliação, obriga os professores a defini-los previamente à administração da avaliação. Também, Gréhaigne e Godbout (1998) consideram que a comunicação das expectativas é importante. Para os autores, quando os alunos vão para a prática já devem saber o que estão a tentar alcançar. Defendem que, sem critério de sucesso, os alunos nunca saberão, por eles próprios, se tiveram ou estão a ter sucesso.

De acordo com Martin et al. (2002), a avaliação pode recompensar e ajudar a demonstrar progressos, já que, quem tem confiança vê a avaliação como uma oportunidade para demonstrar competência e melhoria. Para os autores, o facto de os alunos terem uma reduzida participação na sua própria educação gera insatisfação e falta de empenhamento. Concluem que para haver adultos que façam uma gestão activa das suas vidas e pensem autonomamente, é necessário ajudar os alunos a desenvolver competências de auto-regulação.

Quando assumem mais responsabilidade na aprendizagem mostram mais envolvimento e entusiasmo nas aulas (Doolittle, 1986), e como participantes activos no seu processo de desenvolvimento e na sua equipa aprendem melhor (Hubball, 2004). Como mostra o estudo de Nazario (2005), quando os alunos estão envolvidos na auto-avaliação assumem responsabilidade pelo desenvolvimento das suas capacidades motoras.

Apoiando-se noutros autores, Wood (2003) explica que a avaliação não só deve examinar o produto de aprendizagem, o trabalho deve ser apresentado publicamente, como também, examinar o processo, sendo que na EF uma das características da avaliação autêntica é a melhoria, participação e esforço dos alunos contarem.

Neste quadro, em coerência com todos os aspectos analisados neste ponto, também na EF devem ser consideradas as características tidas como importantes para que a avaliação contribua para uma educação que dê resposta aos desafios do nosso tempo, ou seja, uma avaliação visando competências de nível superior, centrada em estratégias metacognitivas que os alunos utilizam como participantes activos, que valoriza os aspectos de natureza sócio-afectiva e o trabalho colaborativo, e que está integrada no processo de ensino-aprendizagem.

### 1.5. Discussão da especificidade dos JDC em EF

Em Portugal, a importância que o jogo tem no currículo dos alunos em EF, é evidenciada pelos autores dos programas, Bom et al. (1990), que ao explicar a estrutura do curso de EF do 1º ao 12º ano e a sua composição curricular destacam no 1º ciclo, o bloco de Jogos. Este bloco, contém objectivos de aprendizagem dos fundamentos do Futebol e Voleibol, e também, jogos de preparação que asseguram as aprendizagens que antecedem a introdução do Basquetebol e do Andebol no currículo. Aliás, o reportório motor de base destes jogos surge desde o 1º ano no bloco sincrético de Perícias e Manipulações. Já no 2º e no 3º ciclo, os jogos aparecem nas subáreas de Jogos de Raqueta, Jogos Tradicionais e Populares e, sobretudo, na de JDC. No Ensino Secundário, depois do 10º ano que apresenta predominantemente um carácter de revisões sobre o que foi aprendido, surge o regime de opções no 11º e 12º ano, que corresponde a preferências de aperfeiçoamento dos alunos e que, em coerência com um currículo articulado verticalmente, obriga à escolha de dois JDC.

Por outro lado, também as Normas de Referência para o Sucesso em Educação Física [NRSEF] (Jacinto et al., 2001), surgidas na sequência da revisão curricular, revelam a obrigatoriedade do domínio de níveis do programa, confirmando a importância dos alunos demonstrarem competências na subárea dos JDC<sup>12</sup>.

Assim, para melhor compreender aspectos que a avaliação autêntica nos JDC deve contemplar importa perceber o que são estes jogos e o que significa jogar bem. Tais formulações têm importância na forma como se entende o papel da tática e da técnica no desempenho. Em matéria de avaliação, a forma como nos situemos em relação à ligação entre a técnica e a tática pode determinar a coerência das abordagens de avaliação e é factor decisivo para discutir quaisquer opções metodológicas dessas abordagens.

#### *Definição de Jogo Desportivo Colectivo e de "Jogar Bem"*

Um jogo define-se como uma actividade física usando um objecto que é jogado em sociedade, como o Futebol, o Ténis, o Golfe, ou o Softbol (Hopper & Bell, 2001). Para Almond (1986a), nos jogos a luta pelo domínio territorial é decidida por um sistema de pontuação que simboliza a extensão da vitória. Para o autor essa luta inclui estratégia significativa e aspectos técnicos em que a antecipação é primordial, sendo o código de regras o elemento que assegura que ambas as equipas ou individuais se encontram numa base de igualdade.

Como se pode verificar tais definições abarcam todo o tipo de jogos, não só os que são jogados em equipa, como também os que implicam uma prestação individual ou em

---

<sup>12</sup> De acordo com as NRSEF, no 3º ciclo, os alunos têm de evidenciar competências de 1 nível elementar e de 6 níveis introdução e, no Secundário, de 3 níveis elementar e de 3 níveis introdução. Em ambos os casos, 3º ciclo e Secundário, é obrigatório considerarem-se 2 JDC.



pares. Para melhor compreender os jogos que interessam ao âmbito do nosso estudo, importa sistematizar, em função das suas características, a forma como os jogos podem ser organizados.

Na sua leitura da matéria publicada, Hopper e Bell (2001) destacam, a partir de algumas das classificações mais populares, quatro tipos de jogos:

Os jogos com alvo; os jogos com rede/parede; os jogos com batimento; os jogos territoriais.

Na mesma linha Almond (1986b) classificou-os da seguinte forma:

Os jogos com alvo; os jogos com rede/parede; os jogos de campo/marcação em corrida; os jogos de invasão.

Quadro 2. Classificação de jogos. Adaptado de Almond (1986b).

<b>INVASÃO</b> Podem ter alvo focado (baliza ou cesto) ou alvo aberto (linha)	Mão	Basquetebol Andebol Corfebol Pólo Aquático
	Pé	Futebol Futebol Americano Rugby
	Stick	Hóquei Hóquei em patins Hóquei no gelo Pólo
<b>REDE/PAREDE</b> Campo dividido ou partilhado	Rede/raquete	Badmington Ténis Ténis de Mesa
	Rede/mão	Voleibol
	Parede	Squash Pelota
<b>CAMPO/MARCAÇÃO EM CORRIDA</b>		Basebol Softebol Cricket
<b>ALVO</b> Podem ser com oposição ou não.		Golfe Curling Bowling Bilhar Snooker

A partir da literatura publicada Graça e Mesquita (2007), afirmam que é possível estabelecer critérios para colocar os jogos nos currículos baseados no sistema de classificação dos jogos. Há conceitos como compreensão estratégica e consciência táctica que podem ser ensinados porque se transferem entre jogos (Hopper & Bell, 2001). De acordo com Mandigo, Butler e Hopper (2007), os alunos estão capazes de participar em vários jogos se dominarem aspectos relacionados com cada categoria, tais como, habilidades fundamentais, problemas tácticos e regras principais. Para os autores, é o que acontece nos jogos de invasão se compreenderem o que a manutenção da posse de bola

significa, destacando que, várias soluções táticas como o uso de passes curtos, protecção da bola e apoio ao portador da bola, são transferíveis em jogos similares como basquetebol, andebol, futebol e hóquei.

Na mesma linha, Gréhaigne et al. (2005) referem que para classificar os jogos deve ter-se em conta o tipo de alvo, que pode influenciar a técnica e tática, e a maneira como se interage com o projectil. Os mesmos autores consideram que as regras principais dão ao jogo o seu carácter essencial, estabelecendo a identidade de cada jogo. Sublinham que as regras devem, inclusivamente, ser consideradas o elemento chave do jogo de um ponto de vista ético.

Os jogos que nos interessa analisar e estudar são os jogos de equipa. Bayer (1985) descreve estes jogos como aqueles em que existe uma bola, pela qual lutam as equipas em confronto num terreno de jogo, existindo um alvo a atacar e outro a defender, colegas de equipa para cooperar e adversários cuja oposição se quer vencer, enquadrados por um conjunto de regras que devem respeitar-se.

Téodorescu (1977) utilizou a designação, JDC, para referir-se a este tipo de jogos. Esta designação é importante porque permite distinguir outros desportos colectivos como estafetas ou remo.

Quando nos referimos a JDC estamos então a considerar jogos como o Andebol, o Futebol, o Basquetebol, o Voleibol ou o Basebol. Mais concretamente, nos JDC de invasão, que se distinguem do Voleibol ou do Basebol, para que o objectivo de pontuar se concretize, uma equipa invade o campo da equipa adversária (Oslin et al., 1998; Garganta, 1998), realiza as habilidades ditas abertas em ambiente de grande incerteza e imprevisibilidade (Konzag, 1991), luta pela posse da bola e utiliza predominantemente trajectórias de circulação da bola (Garganta, 1998), com cada um dos seus elementos tentando ultrapassar ou evitar obstáculos móveis, abrandando ou parando a progressão dos adversários, procurando produzir incerteza nos opositores e reduzi-la para os colegas de equipa e aceitando o projecto colectivo, dando com as suas acções o melhor pelo grupo (Gréhaigne et al., 2005).

Na sua perspectiva desta área de estudo, Oslin et al. (1998) constataam que os aspectos ofensivos e defensivos do jogo são similares no futebol, basquetebol e hóquei. Tomando como referência outros autores, trata-se segundo Gréhaigne et al. (2005) e Gréhaigne et al. (2001), de cada jogador cooperar com os colegas para melhor se oporem aos adversários, revelando níveis de prontidão para defender enquanto atacam e para atacar enquanto defendem.

Assim, interessa-nos neste tipo de jogos, perceber quem são os bons jogadores. Quem é que tem um desempenho individual que influencia o rendimento colectivo, ou seja, quem coloca a sua equipa mais perto da vitória?

De acordo com Gréhaigne et al. (2005) e, também, Gréhaigne et al. (2001) os bons jogadores são aqueles que durante o jogo, com frequência e no momento apropriado, escolhem e executam bem as acções adequadas.

“Performance in team sports appears to be determined by the most appropriate choice among the various solutions at the players’ disposal and by the speed of this decision-making” (Gréhaigne et al., 2005, p. 28). Para os autores, frequentemente várias respostas permitem resolver um problema, a mais simples é normalmente a mais adequada e segura porque “é a mais económica sob um ponto de vista mecânico, energético, emocional e informacional” (p. 60). Para Martin & Gaskin (2004), a velocidade como a decisão é tomada é, precisamente, uma das coisas que mais diferencia o jogador experiente do novato.

A «força de rácio»<sup>13</sup> e o potencial do jogador para ganhar vantagem numa situação específica, que têm a ver com a configuração de jogo, são dois aspectos determinantes para analisar o processo de jogo (Gréhaigne, Caty & Godbout, 2010).

Na perspectiva de Gréhaigne et al. (2005), um bom jogador do ponto de vista táctico consegue avaliar a presença numa situação potencial em que se pode adquirir vantagem e percebe, em cada momento, o que é bom para a sua equipa e mau para o adversário. Para os autores, um bom jogador exerce pressão no jogo, influenciando a evolução da configuração do jogo, que é a noção relativa ao posicionamento dos jogadores de ambas as equipas em relação à posse e localização da bola, ou seja, um bom jogador cria uma configuração de jogo que é vantajosa.

#### *Técnica versus Táctica: Uma disputa sem sentido*

Se tomarmos em consideração a definição do que é jogar bem e as características essenciais dos jogos colectivos, a dicotomia entre a técnica e a táctica fica sem sentido. De facto, não há uma resposta simples para a complexidade do ensino dos jogos e esta disputa só faz perder tempo (Hopper, 2002).

Revendo a literatura, Graça e Mesquita (2007) consideram essa dicotomia uma questão falaciosa porque o desempenho no jogo depende das duas e os próprios constrangimentos de natureza técnica limitam a tomada de decisão.

Passos, Batalau e Gonçalves (2006) consideram que as circunstâncias do envolvimento determinam a existência de problemas de natureza táctica e sublinham que a

---

<sup>13</sup> Segundo os autores, a noção de «força de rácio» corresponde à dinâmica da movimentação de uma para a outra baliza durante cada sequência de jogo em que na luta pela posse de bola se confrontam um grupo de jogadores com outro grupo de jogadores.

técnica é, precisamente, a acção coordenada que dá resposta a um problema táctico. Para Júlio (2006), é a táctica que justifica a necessidade das técnicas e lhes determina o contexto. Em síntese, as técnicas são uma ferramenta da táctica, são acções e, também, elementos que integram acções mais complexas (Mahlo, 1969/n. d.).

Nos JDC, o jogador tecnicamente competente relaciona-se com bola utilizando preferencialmente o controlo quinestésico, ficando o olhar liberto para observar o jogo e, assim, poder tomar melhores decisões (Mesquita, 2000). Se as técnicas forem adequadamente dominadas e exercidas não exigem atenção para os pormenores de execução, ou seja, para a sequência ajustada de elementos da coordenação motora, ajustamento desses elementos entre si e ao contexto da tarefa. É isto que significa que a aprendizagem culmina no automatismo dos pormenores de execução, ficando a consciência disponível para a táctica (Mahlo, 1969/n. d.).

Mas a capacidade de controlar e direccionar o objecto não é o único indicador de habilidade, também, outros aspectos críticos de habilidade no jogo são importantes, como apoiar colegas de equipa com bola abrindo linhas de passe com ângulos e distâncias correctos e na defesa, fazer cobertura a colegas com ângulos e distâncias correctos (Harvey, 2006).

Por outro lado, durante o processo de ensino e aprendizagem interessa considerar os momentos em que a atenção deve estar mais focada em cada um dos aspectos e como se deve estabelecer essa relação.

Considerando os trabalhos sobre este tema, Antunes e Sampaio (2003) defendem que devem estabelecer-se relações mais produtivas entre a técnica e a táctica, segundo o princípio que, o momento mais apropriado para a prática da técnica está dependente do momento em que o seu valor é reconhecido no contexto de jogo. Segundo Mesquita (2000), não faz sentido saber do ponto de vista técnico mais do aquilo que o jogo exige do ponto de vista táctico, ou seja, a lógica da didáctica deve estar em consonância com a lógica do jogo. Para a autora, embora seja difícil, no treino das habilidades, esta relação de compromisso deve ser respeitada.

De acordo com Gréhaigne et al. (2005), “the educational setting should not oppose technical learning and tactical learning but rather should articulate them in a tactical matrix of technique” (p. 20). Para os autores, o jogo real deve ser jogado desde início e a estratégia do professor deve conduzir à compreensão e aplicação de princípios tácticos, através da resolução de problemas tácticos e estratégicos, e logo que sinta necessidade, promover a aprendizagem de habilidades técnicas adequadas. Para Mesquita (2000), como nas situações de jogo há uma interacção entre a selecção e a execução da resposta motora, os ensinamentos da técnica e da táctica devem estar relacionados. Assim, para a autora, também as

tarefas motoras devem exigir, igualmente, capacidade de decisão e capacidade de concretização.

“It is too easy for teachers to focus on content, believing they are teaching tactics or techniques, when in reality they are covering material but not engaging the learner” (Hopper, 2002, p. 48). O autor defende que a habilidade do professor está em mudar da focagem técnica para a tática de acordo com as necessidades dos alunos, adaptando-se às suas variações de habilidade e conhecimento.

“The problem is therefore to reduce the whole complexity of the game to the smallest number of elements, articulated to a highly logical functional unit” (Gréhaigne et al., 2005, p. 21). Para os autores, não havendo praticamente duas situações de jogo iguais e sendo quase impossível recriar todas as situações, é no entanto possível, consideradas as suas características, padronizar em categorias o tipo de situações que constituem muitas unidades tácticas separadas.

Por outro lado, estes autores também reconhecem que o potencial atlético é um factor limitativo para a actividade e que há uma interdependência entre a tática, a técnica e o potencial atlético. Concluem que, como a execução motora de uma escolha táctica requer energia física, a fadiga pode, também, afectar a lucidez das escolhas e, não só, a qualidade da execução.

## **1.6. Modelos de ensino baseados numa aprendizagem autêntica dos jogos**

Nos últimos trinta anos, o processo de aprendizagem dos JDC passou a ter como referência central o próprio jogo. A capacidade de jogar e compreender o jogo, resolvendo os problemas que ele coloca, substituiu o ensino isolado e descontextualizado das técnicas (Barreto et al., 1980; Graça & Mesquita, 2007). Esta viragem manifesta-se em vários modelos de ensino dos jogos e têm influenciado a prática dos professores no sentido de tornar a aprendizagem mais autêntica.

No entanto, num passado já distante, parece que esta preocupação já dominou o interesse dos investigadores. Num estudo que realizou, Rodgers (1937) teve como objectivo descobrir quais dos três métodos de ensino, em jogos de equipa como o Futebol, o Voleibol e o Basebol, resultou em desempenhos mais competentes no jogo, proporcionou maior conhecimento das regras e foi mais divertido para aprender a jogar:

- 1 – Jogar o jogo sem treinar as técnicas.
- 2 – Exercitar as habilidades do jogo em 90% do tempo e jogar o jogo em 10%.
- 3 – Exercitar as técnicas relacionando-as com a necessidade de melhorar a habilidade técnica enquanto se joga o jogo.

Nesse estudo participaram oito professores de EF e 600 alunos rapazes de sete escolas públicas de Cleveland, e seis professores sem especialidade em EF e 500 alunos rapazes de seis escolas de Lakewood. A amostra não foi randomizada, mas não parece haver evidência de enviesamentos nas escolhas feitas.

Segundo a autora, os resultados do estudo parecem claros. Em ambos os casos, com professores especialistas e sem especialidade em EF, os alunos tornaram-se mais competentes em futebol, voleibol e basebol, utilizando o terceiro método em comparação com os outros dois.

Também há uma significativa diferença quando são avaliados os dados recolhidos sobre conhecimento das regras.

Os alunos também julgam ser este método o mais divertido para aprender a jogar e os professores também o avaliam como o mais eficiente e eficaz, acreditando que assim os alunos têm mais prazer em jogar. Os dados mostram a superioridade dos professores de EF em relação aos outros na forma de ensinar as competências dos jogos de equipa.

Ora, esta preocupação com uma aprendizagem mais contextualizada pode ser constatada na literatura publicada onde se podem encontrar vários modelos de ensino dos jogos, e que iremos aprofundar.

#### *O Ensino dos Jogos para a Compreensão - Teaching Games for Understanding*

O modelo de ensino dos jogos para a compreensão foi sistematizado no seu esboço essencial por Bunker e Thorpe (1982).

O modelo desenvolve-se em seis passos:

1- Fomentar o conhecimento e compreensão do jogo através de uma forma modificada do jogo formal, que reflecta os conhecimentos, técnicas e competências de uma categoria de jogos e seja adequada ao nível dos alunos.

2- Desenvolver a apreciação do jogo. Ao apreciar as regras, o que se pode ou não fazer, qual o sistema de pontuação, etc., percebe-se o significado de como cada jogo é jogado.

3- Desenvolvimento do conhecimento táctico através da consciencialização dos principais problemas tácticos do jogo e compreensão das tácticas elementares defensivas e ofensivas que permitem ganhar vantagem sobre o adversário.

4- Contextualizar a tomada de decisão. Tomando como referência o que aprenderam na apreciação do jogo e no conhecimento táctico as situações de jogo são interpretadas de forma a escolher o que fazer e como executar.

5- Executar. Ao tomar consciência de como as habilidades técnicas são importantes para resolver problemas do jogo, melhora-se e refina-se o seu domínio, em contexto.

6- Enfatizar o desempenho. Aplicam-se os passos anteriores em situação de prestação, podendo inclusive, introduzir-se um novo jogo ou uma forma modificada mais complexa.

As formas de jogo adaptadas obedecem a quatro princípios:

1- Selecção do tipo de jogo. Utilizar uma amostra de jogos da mesma categoria para melhor compreender as similaridades entre eles.

2- Modificação por representação. Modificar os jogos de maneira a conter a mesma estrutura táctica do jogo formal, basicamente utilizando formas reduzidas representativas do jogo adulto.

3- Modificação por exagero. Exagerar certos aspectos do jogo, manipulando as regras, o espaço e o tempo, para confrontar os alunos com problemas tácticos que levem à descoberta de soluções tácticas básicas.

4- Complexidade táctica. Ajustar o nível de complexidade para que respeitando a capacidade do aluno se desafie a sua capacidade de compreensão e actuação no jogo. Tomar em consideração que os jogos de alvo são os menos complexos, seguidos pelos jogos de rede e os jogos de batimento, e que os jogos de invasão são os mais complexos.

Para Bunker e Thorpe (1986a,b,c) e Thorpe e Bunker (1986), a tomada de decisão baseia-se no conhecimento táctico e precede a resposta dos factores de execução. Se os alunos não compreenderem o jogo, não conseguem seleccionar a resposta adequada.

Estes autores não aceitam que a táctica deva esperar pelo desenvolvimento das habilidades, que devem ser ensinadas e treinadas, quando forem necessárias, para aumentar o conhecimento táctico e para ajudar a tomar melhores decisões. Referem, também, que as habilidades do jogo devem ser ensinadas através dos princípios de jogo. A ênfase deve estar nos princípios e problemas e não nas habilidades e técnicas. Pretende-se, assim, progredir do porquê para o como em vez do contrário, ou seja, das tácticas para as habilidades.

Na sua revisão desta matéria, Harvey (2006) considera que os princípios pedagógicos do TGfU estão em oposição à abordagem tradicional de ensino dos jogos centrada na técnica. Para Hopper (2002), o TGfU não é uma questão de tácticas ou técnicas, mas de progressão das tácticas para as técnicas e afirma que “não há nada de errado em ensinar técnicas quando os alunos estão preparados para isso” (p. 44). Harvey (2006) é um autor, entre outros, representativo da ideia que o TGfU tem o potencial para facilitar as aprendizagens das habilidades e movimentos com bola, mas também de movimentos sem bola

Ao ligar as experiências nas aulas às suas concepções anteriores sobre os jogos, este modelo, proporciona aos alunos uma aprendizagem mais autêntica e significativa (Kirk

& McPhail, 2002). O TGfU proporciona uma visão mais holística do jogo porque o aluno está no centro do processo de aprendizagem (Harvey, 2006). É aceite a perspectiva construtivista baseada na ideia que é ao aluno que compete construir activamente as suas próprias aprendizagens e reconhece-se que um estilo de ensino de descoberta guiada está em consonância com o modelo (Graça et al., 2003; Graça, 2003; Graça & Mesquita, 2007).

Mais recentemente, assumindo a perspectiva construtivista, Gréhaigne et al. (2010) propõem a designação alternativa «aprender para a compreensão». Para os autores, a questão não está em os alunos dominarem o que está pré-estabelecido mas em construir soluções. Partem do princípio que a aprendizagem dos jogos envolve um processo cognitivo da parte dos alunos e que a compreensão do jogo tem o papel maior neste processo.

Este modelo de ensino pretende que os alunos pratiquem um jogo de melhor qualidade, compreendendo-o, estando mais motivados e envolvidos, e conjugando o treino da técnica com as necessidades que o jogo vai revelando (Graça, 2003). Mandigo e Holt (2002) propõem a inclusão do “optimal challenge” (p. 3)<sup>14</sup> com actividades equilibradas que evitem o aborrecimento ou a ansiedade. Para Holt, Strean e Bengoechea (2002), o modelo permite um aumento do nível da motivação dos alunos e pretendem, por isso, a incorporação de mais componentes afectivas no processo pedagógico de ensino, considerando, por um lado, que o domínio afectivo não tem sido tão analisado como os domínios psicomotor e cognitivo, e por outro lado, que é importante perceber como é que o aluno está motivado para poder continuar a participar. Os autores consideram que “o aluno tem estado no centro do modelo TGfU, mas a experiência do aluno não tem estado no centro do debate académico” (p. 164).

Por outro lado, na sua leitura da matéria publicada, Hopper (2002) afirma que muitos investigadores compararam a eficácia das abordagens técnica e táctica, com desenhos de investigação experimental que as opunham. Os resultados destes estudos foram inconclusivos e não provaram a superioridade de nenhuma das abordagens (Hopper, 2002; Graça & Mesquita, 2007).

Martin e Gaskin (2004), são autores entre outros, defendendo que a discussão sobre os méritos de uma abordagem em oposição à outra, como se de extremos opostos se tratassem, é limitativa porque a aprendizagem é complexa e variada e não se reduz a um modelo linear. Também Graça e Mesquita (2007) consideram que nas investigações realizadas se introduziu um artificialismo que levou a instrução a não utilizar elementos da outra abordagem, quando, na verdade, o TGfU nunca negou a necessidade do trabalho técnico. Justamente, o facto de alguns estudos terem desenhos de investigação diferentes,

---

<sup>14</sup> Segundo os autores, os professores devem encorajar os seus alunos a escolher o nível de prática adequado. Um nível que seja, nem demasiado fácil, nem demasiado difícil.



torna-os inconclusivos e impede que se façam comparações válidas dos seus resultados (Harvey, 2006).

Na sua revisão da matéria publicada, Graça e Mesquita (2007), concluem que não faz sentido comparar abordagens que se distinguem quer nos objectivos a atingir, quer nos processos que utilizam. Segundo os autores, a pesquisa actual já se debruça sobre o funcionamento e efeitos produzidos pelas diferentes abordagens e obstáculos que ocorrem no processo de aprendizagem. Apoiando-se noutros autores, Harvey (2006), considera que a pesquisa actual procura orientar-se no sentido de usar ambientes mais contextualizados e eliminar desenhos comparativos, e é acompanhada nesse esforço por teorias da aprendizagem como a teoria construtivista, a dos constrangimentos, a da aprendizagem situada, e as que estão associadas ao afecto e à emoção.

De acordo com Holt et al. (2002), é necessário combinar as duas abordagens. Para os autores a questão que se coloca é sobre o momento da introdução da táctica ou, das técnicas.

#### *Tactical Games*

Os TG (Griffin et al., 1997; Mitchell et al., 2003) são uma extensão do TGfU com princípios muito semelhantes aos do modelo original. São propostos vários níveis de complexidade táctica e um quadro mais autêntico para avaliar o desempenho no jogo. É um modelo que parte de um jogo modificado que revela um problema táctico. Esse problema passa a constituir o foco da instrução, com colocação de questões que permitam compreender como se resolve o problema, desenvolvendo a consciência táctica, e com práticas situadas que permitam treinar habilidades essenciais ou movimentos para resolver os problemas tácticos. Aquilo que se praticou em situações autênticas tem oportunidade de ser aplicado no jogo final.

É de acordo com estes pressupostos que, por um lado, Oslin et al. (1998) constataam que o desempenho em jogo recebe o contributo da habilidade, mas sem conhecimento táctico e compreensão é difícil ou quase impossível jogar o jogo e, por outro lado, Dyson et al. (2004) consideram ser o questionamento uma habilidade crítica de ensino que orienta os alunos na procura de soluções para o problema táctico. Nesta perspectiva, os autores defendem que, tanto quanto possível, os jogos devem ter poucas habilidades, poucas regras e poucos jogadores.

#### *Aprendizagem Situada*

Kirk e McPhail (2002) propuseram uma revisão do modelo adoptando a teoria da aprendizagem situada. O conceito principal é tornar as experiências de aprendizagem autênticas, significativas e credíveis e, alinhá-las com as concepções culturais que os alunos trazem para a escola. Para os autores, os alunos querem ver as tarefas de aprendizagem

como autênticas e em ligação com o jogo, sendo por isso importante que o jogo praticado na escola esteja relacionado com aquele que eles vêem ao vivo ou na televisão. Brooker, Kirk, Braiuka e Bransgrove (2000) concluem do seu estudo que os alunos podem ficar frustrados se nas aulas de Educação Física for muito modificada a versão do jogo adulto que eles bem conhecem, já que os alunos não conhecem o esforço dos jogadores e das equipas para conseguirem jogar assim.

Para haver uma ligação mais conseguida entre o conhecimento declarativo e o conhecimento processual, esta proposta substitui os elementos anteriores «apreciação do jogo» e «consciência táctica» pelos elementos «conceito de jogo» e «pensamento estratégico» (Graça & Mesquita, 2007).

Revendo a bibliografia, Harvey (2006) considera que pode ser eficaz a sugestão de, a partir do TGfU, utilizar uma abordagem de ensino que tenha como referência a prática. Para Dyson et al. (2004) a teoria da aprendizagem situada é uma componente da teoria construtivista da aprendizagem em EF. Na sua visão o currículo está centrado na aprendizagem, dando oportunidade a que cada aluno ajude os outros alunos a aprender.

Gréhaigne et al. (2005) consideram que há três elementos chave na perspectiva construtivista do processo de aprendizagem dos JDC que são a observação, o pensamento crítico e a transformação. Para os autores, o aluno recolhe informações baseado nas observações pessoais porque é ele que está envolvido na acção e sublinham, também, a importância da reflexão na aprendizagem da resolução de problemas e na construção de conhecimento pessoal.

Considerando os estudos publicados sobre o ambiente cooperativo nas aulas, Dyson et al. (2004) referem que os alunos constroem conhecimento através da interacção social com os seus pares, esperando-se que o professor assuma o papel de facilitador das aprendizagens. Nesta linha, também Harvey (2006), é um autor entre outros, representativo da ideia que o papel de um professor construtivista é ajudar os alunos a construir o conhecimento relacionado com os problemas tácticos que são foco na aula.

Fazendo o ponto da discussão teórica, Dyson et al. (2004) explicam que um conceito chave na aprendizagem situada é a participação numa comunidade de prática, ou seja, num grupo em que os seus elementos partilham as práticas numa determinada área da vida.

#### *Educação Desportiva – Sports Education*

O modelo de SE desenvolvido por Siedentop (1994) toma em consideração as principais características do desporto, que são o seu grande atractivo. A partir da noção de autenticidade, promove experiências desportivas autênticas que estão em oposição com as

abordagens tradicionais e descontextualizadas do ensino dos JDC nas escolas. O objectivo é criar pessoas competentes, cultas e entusiastas

Para o autor, há seis características do desporto institucionalizado que tornam essa experiência desportiva autêntica: As épocas desportivas contrastam com as unidades didácticas pequenas que não dão tempo a que o que é típico de um desporto particular seja transmitido e aprendido. Os jogadores estão filiados em equipas, tomam parte activa no processo de ensino-aprendizagem assumindo posições de liderança e responsabilidades em vários papéis como jogadores, árbitros, treinadores, dirigentes, e outros. Há um quadro competitivo formal, cujo objectivo é ganhar jogos, normalmente modificados e reduzidos, mostrando bom comportamento desportivo e organização. Os torneios podem nem chegar ao jogo formal. As equipas são formadas de forma a garantir equilíbrio na competição e colaboração na aprendizagem. Há um evento culminante. Há registos estatísticos. Há uma atmosfera festiva.

Dyson et al. (2004) consideram que a educação desportiva tem o potencial para fazer a ligação dos jovens a uma comunidade de prática a partir do momento que seja reorganizada e enquadrada de forma educativa.

De acordo com Graça e Mesquita (2007), este modelo preconiza uma avaliação autêntica centrada, prioritariamente, na capacidade de jogar. Para os autores, ao avaliar-se desempenhos contextualizados, alinha-se a instrução com a avaliação sem suspender o processo de aprendizagem.

Todavia, num estudo realizado por McCaughtry, Sofo, Rovegno e Curtner-Smith (2004), em que participaram professores novatos, identificaram-se um conjunto de resistências, que surgem na aplicação deste modelo e que correspondem a crenças instaladas, a saber:

- 1 – O modelo é muito trabalhoso para um professor novato;
- 2 – O modelo prejudica a aprendizagem das habilidades motoras;
- 3 – Os alunos não querem fazer as outras tarefas como arbitrar e treinar.

Segundo os autores, como os professores não acreditam o suficiente, a instrução também não é suficiente e adequada. Muitos pensam, erradamente, que o ensino das habilidades não faz parte do modelo, que é só elaborar torneios e que as habilidades se ensinam antes. Também, não compreendem que há outras formas de aprendizagem com valor educativo como o ambiente desportivo, o clima festivo, o bater de records, e outros. Os autores defendem que treinar, arbitrar e fazer estatísticas são tarefas complexas que exigem tempo para ensinar e para aprender, e sugerem que as habilidades possam ser ensinadas na pré-época, nos aquecimentos da época regular e de outras maneiras criativas.

### *Confluência do TGfU e do SE*

Vários autores são a favor da conjugação dos dois modelos (Kirk & McPhail, 2002; Dyson et al., 2004; Graça et al., 2003).

Graça e Mesquita (2007) consideram que ambos dão importância à tática, sendo a componente cognitiva mais acentuada no TGfU.

Nos respectivos trabalhos de discussão da bibliografia, Harvey (2006) e Graça e Mesquita (2007) constataam que a combinação das duas abordagens não só não reduz o currículo, como dá mais ênfase ao professor na condução dos processos, convocando-o a dominar mais conteúdo e conhecimento pedagógico. Ambas privilegiam os jogos reduzidos e a contextualização das situações de jogo.

No entanto, para Graça (2003), professores mal preparados para utilizar o TGfU poderão não ajudar os alunos a ler taticamente o jogo, se transformarem a instrução em questionamento abstracto ou irrelevante, e professores mal preparados no SE poderão deixar os alunos desapoitados, transformando a aula num simples recreio, apesar de organizada e supervisionada.

Dyson et al. (2004) propõem um modelo que congrega o SE, TG e a Aprendizagem Cooperativa, sendo a aprendizagem situada o referencial teórico de ligação entre eles. Os modelos têm princípios pedagógicos similares:

- A aprendizagem está centrada na participação activa dos alunos envolvendo processos de tomada de decisão, interacção social e compreensão cognitiva;
- O professor facilita as actividades de aprendizagem delegando responsabilidades aos alunos;
- Pretende-se uma educação holística nos domínios social, físico e cognitivo;
- Trabalha-se em pequenas comunidades de prática constituídas por poucos elementos que dependem positivamente uns dos outros.

Esta intercepção dos três modelos permite fornecer uma estrutura para que os alunos organizem o seu próprio torneio, com pequenas equipas onde se definem claramente os papéis e as responsabilidades, e se ensine os alunos a cooperar de forma a competir mais eficazmente, trabalhando como equipa contra outra, trazendo para o primeiro plano aspectos da tomada de decisão e da competência para resolver problemas.

### *Modelo de Competências*

O modelo de competências nos jogos de invasão (Graça et al., 2003) é claramente influenciado pelo modelo dos jogos para a compreensão e pelo modelo de educação desportiva. As formas modificadas e simplificadas dos jogos desportivos colectivos de invasão são o contexto escolhido para aprender.

Pretende-se que os alunos sejam competentes como jogadores, trabalhando em equipa, tomando decisões adequadas e executando a solução escolhida com eficácia e eficiência, mas também que sejam competentes em funções de apoio e orientação como capitão de equipa, árbitro, juiz de mesa, observador e treinador.

A aprendizagem desenvolve-se articulando três formas operacionais de trabalho com os alunos, decompondo-as e recompondo-as em sucessão e alternadamente, as formas básicas de jogo, as formas parciais de jogo e as tarefas baseadas no jogo:

- As formas básicas do jogo, com um conceito idêntico ao jogo completo para serem vistas como autênticas pelos alunos, modificam o jogo formal adaptando-o ao nível dos alunos para que se dê um salto para um nível de complexidade mais exigente;

- As formas parciais de jogo, são situações mais simplificadas, que põem em evidência partes do jogo acentuando o problema que se quer ver resolvido, mantendo o objectivo do jogo, as relações de cooperação e oposição, embora não se obrigue a converter o ataque em defesa e vice-versa nem à manutenção de mesmo número de jogadores para cada equipa;

- As tarefas baseadas no jogo, que mantêm como referência a situação de jogo em foco, permitem que em contexto muito simplificado, o que torna as escolhas óbvias, se melhore a execução das habilidades do jogo.

Preconiza-se a utilização de regras facilitadoras e de regras de acção que permitam aos alunos o controlo do corpo, o controlo da bola e o controlo do jogo, mantendo-se o equilíbrio dinâmico entre pessoa, tarefa e envolvimento.

Preconiza-se que a avaliação dos alunos seja autêntica. Uma avaliação em contexto de jogo, nas formas básicas, que avalie o desempenho, quer como jogadores, quer nas funções de apoio e orientação. Sugere-se a utilização do GPAI e do TSAP como instrumentos de avaliação autêntica do jogo.

#### *Abordagem Ecológica*

Assenta na perspectiva ecológica da aprendizagem que, segundo o modelo de Newell (1986), considera a existência e interacção de três categorias de constrangimentos, os do sujeito ao nível funcional e estrutural, os do meio envolvente e os da tarefa. Manipular os constrangimentos permite o envolvimento dos alunos na exploração e descoberta de soluções para os movimentos e está de acordo com a perspectiva da pedagogia não linear (Chow et al., 2007).

Do ponto de vista das dinâmicas ecológicas, são as interacções contínuas entre os alunos e o ambiente que geram as informações que permitem aprender em ambientes dinâmicos. A maior implicação das características dos sistemas não lineares na aprendizagem, é que pequenas mudanças nas tarefas práticas podem resultar em

mudanças significativas no comportamento dos alunos (Chow, Davids, Hristovski, Araújo & Passos, 2011). Para os autores, o comportamento emerge, justamente, desta adaptação aos constrangimentos, sendo os professores facilitadores que orientam os alunos a encontrar as soluções, através da exploração das actividades.

Segundo Passos et al. (2006), a abordagem ecológica está baseada na manipulação dos constrangimentos. É a interacção mútua entre os constrangimentos do sujeito, do envolvimento e da tarefa gera campos de informação para que o atleta empreenda acções e tome decisões. Para os autores, a manipulação dos constrangimentos no processo de treino deve gerar campos de informação semelhantes ao tipo de informação típica da competição. Para Júlio (2006), não são as categorias de constrangimentos que determinam o comportamento, mas sim, a interacção entre jogador e envolvimento, durante o processo relacionado com o cumprimento do objectivo da tarefa.

O objectivo dos alunos não é reproduzir um padrão de movimento ideal mas encontrar a sua própria solução pessoal (Chow et al., 2011). De acordo com Passos et al. (2006), o processo de treino deve permitir uma adaptação do atleta às exigências envolvimento, explorando activamente as possibilidades que se abrem para encontrar o seu próprio caminho. Sublinham que, ao contrário da abordagem cognitivista cuja referência é um modelo ideal de execução, a abordagem ecológica permite que o atleta descubra o seu próprio padrão de execução. Para Júlio (2006), uma das limitações do TGfU é assentar, precisamente, numa abordagem cognitivista da tomada de decisão.

Araújo (2003) põe em causa que a acção táctica seja determinada cognitivamente, que a tomada de decisão preceda a acção. Para este autor “é na sua acção que o jogador decide; decide actuando em função daquilo que o contexto lhe reclama, a sua trajectória é o seu raciocínio” (p.91). Assim, nessa perspectiva, defende que apesar da actividade cognitiva estar imbricada no movimento desportivo, não o precede, embora reconheça haver actividade cognitiva prévia que influencia o movimento. Conclui que não se pode determinar uma hierarquia entre decisão e acção, ou seja, como o contexto está sempre a mudar, a adaptação será impossível se o comportamento já estiver pré-determinado. Para Júlio (2006) explicam-se as decisões e acções dos alunos «no sítio onde eles têm a cabeça», que é o contexto de jogo e não «dentro da sua cabeça».

Em suma, na abordagem ecológica da cognição e da acção, não só os constrangimentos físicos e sociais no comportamento individual não são ignorados, como as características dos indivíduos não são esquecidas (Araújo & Davids, 2009).

#### *Da Actividade Lúdica à Formação Desportiva*

Também, em Portugal, houve uma mudança na perspectiva do ensino dos jogos, procurando-se uma aprendizagem mais contextualizada em sintonia com uma visão cultural

do desporto. Num seminário sobre metodologia do basquetebol (Barreto et al., 1980) partiam do pressuposto que o ensino se baseava em duas perspectivas que se operacionalizavam na prática do seguinte modo:

Numa, decompunha-se o jogo em partes, que depois de treinadas fora da situação de jogo, eram aplicadas no jogo. O ensino estava centrado nas soluções motoras que os alunos utilizavam no decorrer do jogo, numa prática profundamente mecanicista em que o próprio professor substituí o aluno, pensando o jogo por ele.

Na outra, a aprendizagem estava, exclusivamente centrada no jogo. Como jogo, por si só, não é agente transformador do aluno, não se dava oportunidade ao desenvolvimento do repertório técnico.

Partindo destes pressupostos os autores preconizavam um ensino do jogo baseado nos seguintes aspectos:

- Centrar o método de ensino na acção, ou seja, no jogo, pretendendo o conhecimento e domínio prático das acções tácticas e dos princípios que as orientam, havendo uma correspondência entre as acções de jogo e as intenções tácticas, procurando que as técnicas dêem expressão e objectivem as acções tácticas e os princípios do jogo.

- Centrar o método de ensino no aluno, numa participação consciente como sujeito da acção de jogo e que adapta o seu comportamento aos problemas que o jogo coloca. Substitui-se a mecanização pela aprendizagem consciente das acções de jogo, ou seja, a táctica e a técnica, que são permanentemente avaliadas em situação de jogo.

Ao professor estava reservado o importante papel de transmitir o conteúdo e sentido pedagógico do jogo. Pretendia-se que ele conhecesse e dominasse soluções pedagógicas e didácticas que possibilitassem a criação de uma área potencial de desenvolvimento nos alunos. Pretendia-se que tivesse, também, a capacidade de observar e avaliar os resultados da aprendizagem.

### **1.7. Avaliação Autêntica nos JDC – “O Estado da Arte”**

Os modelos que preconizam um ensino mais autêntico dos JDC trouxeram a necessidade duma avaliação autêntica alinhada com essas novas formas de ensinar. Surgem alternativas de avaliação aos tradicionais testes de habilidades sem contextualização.

Apoiando-se noutros autores, Silva (2008) refere que tomaram-se como referência indicadores de natureza técnico-tácticos, procurando, por um lado, objectivar a análise do jogo, e por outro lado, verificar a importância desses indicadores no processo de ensino aprendizagem.

Na sua leitura da matéria publicada, Gréhaigne et al. (2005) referem que investigadores da pedagogia do desporto franceses experimentaram vários procedimentos de avaliação em contexto de jogo, todavia, observações de professores nas suas práticas de avaliação mostram que alguns deles já consideram os aspectos tácticos do jogo nos seus próprios instrumentos de avaliação. Também, segundo Harvey (2006), vários investigadores realizaram estudos prévios sobre o TGfU usando vários protocolos de avaliação.

No entanto, o GPAI e o TSAP, que são instrumentos de avaliação centrados na dinâmica de jogo, revelaram-se na literatura publicada os mais importantes e referenciados.

#### **1.7.1. O Game Performance Assessment Instrument**

O GPAI é um instrumento que está desenvolvido e validado, proporcionando aos professores e investigadores um sistema para observar e avaliar as acções e comportamentos de desempenho no jogo. Avalia esse desempenho ao nível da compreensão táctica e da capacidade para resolver problemas tácticos utilizando adequadamente as habilidades técnicas. Pretende ser um instrumento de observação flexível que pode ser usado para avaliar o desempenho do aluno em situação real ou em videotape.

Os autores do GPAI em conjunto, com especialistas com conhecimento nas quatro categorias de jogo (invasão, rede/parede, batimento e alvo) propuseram sete componentes: base, ajustamento, tomada de decisão, execução de habilidade, apoio, cobertura e marcação. Cada uma destas componentes pode ser avaliada independentemente ou em combinação com outras. Nem todas as sete componentes se aplicam a cada jogo particular. Base e Ajustamento referem-se ao Basebol e as outras cinco a jogos de invasão.

Oslin et al. (1998) definiram o constructo de cada componente da seguinte forma:

Base – Recuperar a melhor posição no jogo entre acções.

Ajustamento – Movimentação ofensiva e defensiva em função do fluxo de jogo.

Tomada de decisão – Escolher adequadamente o que fazer com a bola ou outro projectil.

Execução da habilidade – Executar eficientemente as técnicas seleccionadas.

Apoio – Desmarcação. Movimento sem bola para uma posição onde possa receber um passe.

Cobertura – Apoio defensivo a um jogador para fazer algo com a bola ou se movimentar para a bola.

Marcação – defender um adversário que tem a bola ou pode vir a tê-la.

Algumas componentes como tomada de decisão, execução de habilidade, apoio e ajustamento foram previamente validadas no futebol, basquetebol e voleibol



Baseados na quantidade de acções adequadas e inadequadas em cada categoria de constructo, podem-se contabilizar dois índices globais, o de desempenho no jogo e o de envolvimento no jogo.

O «envolvimento no jogo» pode ser avaliado somando todas as respostas que indiquem envolvimento, incluindo tomadas de decisão inadequadas e execuções ineficientes (Harvey, 2006). Revendo a literatura, Tallir et al. (2004) reconhecem que jogadores fracos, com poucos argumentos técnicos ou que decidem mal, podem ainda assim estar muito envolvidos e ter uma participação significativa no jogo.

«Desempenho no jogo» é uma medida mais precisa e calcula-se somando as pontuações de todas as componentes avaliadas e dividindo-as pelo número de componentes avaliadas (Gréhaigne et al., 2005).

Na sua leitura da matéria publicada, Harvey (2006) conclui que o instrumento tem dois benefícios, pode ser adaptado a vários desportos e permite recolher dados de habilidade com bola e sem bola. Este aspecto valoriza um jogador que abra linhas de passe ou que trabalhe arduamente na defesa para que a equipa ganhe a posse de bola (Harvey, 2007).

Usando o GPAI, Harvey (2004) descobriu que melhoraram substancialmente os aspectos defensivos do desempenho no jogo, incluídos num estudo com jogadores de U19 numa academia de futebol, que jogavam em pequenas unidades de 3 jogadores. O autor atribuiu estas melhorias principalmente ao uso de um currículo desenhado em espiral, mas também às experiências prévias de jogo dos jogadores e à instrução eficaz do treinador, utilizando o TGfU.

No desenho experimental numa investigação, Harvey (2006) utilizou quatro índices: de decisões tomadas, de execução de habilidades, de desempenho no jogo e de envolvimento no jogo. Segundo o autor, calcular o «envolvimento no jogo» e o «desempenho no jogo» fornece aos alunos um grande retrato do seu jogo. Utilizou, também, um sistema de pontuação de 1 a 5<sup>15</sup>, afirmando que o professor tem de criar indicadores para os cinco critérios.

Na sua leitura da matéria publicada, Harvey (2006) considera preferível a utilização de mais do que uma medida de desempenho no jogo porque isso permite avaliar melhor o impacto das intervenções nos diferentes domínios, comportamental, cognitivo e social.

---

<sup>15</sup> O autor considera que este sistema de pontuação, 1-5, é preferível nos jogos de invasão:  
5 Desempenho muito eficaz.  
4 Desempenho eficaz.  
3 Desempenho moderadamente eficaz.  
2 Desempenho fraco.  
1 Desempenho muito fraco.

O autor também afirma que só pelo desenvolvimento destes instrumentos se assegura uma avaliação dos alunos válida e fiável em contexto de aprendizagem autêntica e que, no futuro, é necessário garantir que os instrumentos permitam ensinar e avaliar simultaneamente.

Por outro lado, na tentativa de resolver ou atenuar problemas detectados na aplicação do GPAI, Memmert e Harvey (2008) fizeram um esforço para solucionar cinco deles:

1º - O cálculo do índice de desempenho individual e do índice de desempenho global. Se alguém tem 0 acções apropriadas o seu índice é sempre 0 porque zero sobre qualquer coisa é zero.

2º - Usar o índice de envolvimento no jogo versus o índice de desempenho no jogo. Os índices podem mascarar a verdade. De facto, 1 acção apropriada + 1 acção inapropriada pode ser o mesmo que 5 acções apropriadas + 5 acções inapropriadas.

3º - Fiabilidade dos observadores. Quando há mais do que um só observador, não toma em conta os resultados de todos eles. Um observador pode pontuar o jogador A com 3 e o jogador B com 5 e o outro observador pontuar A com 5 e B com 3. Se o segundo tiver razão, A é melhor que B, mas se houver uma média, então os dois parecem iguais.

4º - Não linearidade. A parte inferior da equação do índice global de desempenho no jogo não é linear. As boas e más respostas são tratadas de maneira diferente.

5º – Utilidade para a acção. É difícil dizer quando uma acção é ou não adequada, especialmente quando se codificam comportamentos sem bola. Diferentes codificadores podem ter diferentes visões quando categorizam os acontecimentos.

#### *A crítica ao GPAI*

Como os índices globais são baseados nas quantidades de acções em cada categoria de constructo, parte-se das partes que constituem cada categoria e da sua soma para se chegar à avaliação do desempenho individual no jogo que, neste caso, constitui o todo. Tal, equivale àquilo que Elias (1997) considera ser, numa perspectiva epistemológica, perder de vista a dependência recíproca das partes, a sua sintonia funcional.

Na tentativa de resolver ou atenuar os cinco problemas detectados na aplicação do GPAI, Memmert e Harvey (2008) fizeram um esforço meritório mas que não ultrapassa tais Imitações. Vejamos:

- Pode-se resolver ou atenuar os problemas detectados com o cálculo do índice de desempenho individual e do índice de desempenho global com uma constante, ou seja, cada jogador começar com a pontuação 10;

- Pode-se separar as acções adequadas das acções inadequadas e não fazer um índice de tudo ou, transportar os dois índices para um índice global como no TSAP. Tal é

possível, mesmo considerando que é complicado o facto das acções inadequadas com bola não serem contabilizadas no índice de “envolvimento no jogo” mas já o serem no índice de “desempenho no jogo”;

- Pode-se responder ao problema da fiabilidade dos observadores, utilizando uma fórmula matemática;<sup>16</sup>

- Pode-se, até, reduzir o problema da não linearidade, começando por tratar da mesma maneira as acções adequadas e inadequadas;

- Já o problema mais difícil de resolver, a utilidade para a acção, não se consegue ultrapassar estreitando apenas os constructos e treinando os observadores, como propõem os autores. A adequação das acções não se consegue observar num quadro de separação das partes componentes que constituem o desempenho no jogo. Esta é a visão equivocada que criticamos, a da separação de componentes que não podem ser separadas.

Será que a tomada de decisão pode ser separada da execução da habilidade? Será que esta separação não poderá conduzir a conclusões equivocadas?

Definidos os constructos, tomada de decisão e execução do passe, e tendo que classificar isoladamente essas componentes do jogo, pode-se garantir que os avaliadores possam estar de acordo que um determinado jogador tomou a decisão certa porque, perante uma linha de passe ofensiva aberta, passou a bola ao colega desmarcado – decisão apropriada e, também possam estar de acordo que apesar de ter tomado a decisão certa, tenha errado o passe, porque a execução técnica não foi a melhor – execução inapropriada. Tal, só é possível porque se avaliam as partes, isoladamente.

É muito duvidoso considerar que alguém tomou a decisão certa, se não conseguiu efectuar o passe com o ângulo, a direcção e a velocidade adequada. Decidir bem, é não efectuar o passe por reconhecer que, eventualmente, não tem os argumentos técnicos necessários para fazer aquela execução. Como explicam Rovegno, Nevett e Babiarz (2001), um jogador pode escolher não realizar uma acção que desempenhe mal, mesmo que saiba ser a melhor solução para a equipa naquela situação de jogo.

De igual modo, alguém que não executou um passe, para um jogador com linha de passe claramente aberta, pode ter tomado a melhor decisão. Em função do ângulo, direcção e velocidade que são necessários para a bola lá chegar, pode representar muita lucidez reconhecer que um colega de equipa, com os recursos que tem, não é capaz de efectuar uma boa recepção.

---

<sup>16</sup> Para resolver o problema da fiabilidade dos observadores os autores preconizam a utilização da “fórmula de Hart” (Memmert & Harvey, 2008, p. 233).

A definição de um bom passe não é independente do jogador que o faz nem do jogador que o recebe, tendo sempre uma decisão intrinsecamente, associada. Seguindo as indicações de Rovegno, Nevett, Brock e Babiarz (2001) uma definição correcta de um «bom passe», consideramos não apenas aquele que é feito apenas com apropriada direcção, trajectória e força, mas principalmente um passe que possa ser recepcionado no sentido da continuidade do jogo, com vantagem para a dinâmica da equipa relativamente à equipa adversária.

A separação entre decisão e execução é completamente artificial, só serve para análise e nunca para apreciar um todo que é uno. Não é razoável reconhecer os pressupostos de análise global do jogo e realizar uma avaliação que fragmenta o objecto, sem considerar a sua unidade dinâmica.

Se aceitamos, como pressupostos, que a separação entre a técnica e a táctica é um mal-entendido (Hopper, 2002) e que introduz uma falsa questão (Graça & Mesquita, 2007), que se deve evitar a separação metodológica e temporal entre as duas (Tavares, 1998), concordando com (Mahlo, 1969/n. d.), segundo o qual a técnica é uma ferramenta da táctica e que as aprendizagens da técnica e da táctica não devem estar em oposição mas articuladas numa matriz táctica da técnica (Gréhaigne et al., 2005), então deve-se avaliar em consonância com essa ideia de ensino autêntico, baseado no contexto de jogo. Se assim não for perde-se a coerência do modelo de análise, separando, em matéria de avaliação, o que não pode ser separado.

Imagine-se, que também introduzíamos componentes relacionadas com as capacidades motoras condicionais para avaliar o desempenho global no jogo, por exemplo, a velocidade. Seria válido avaliar-se isoladamente o deslocamento em corrida, atribuindo-lhe uma pontuação que é somada a outras para obtermos um índice global de desempenho? Seria válido considerar que um jogador deslocando-se muito rapidamente e com elevada frequência, influencia directamente, só por isso, o desempenho global das equipas em jogo?

Nada é mais falacioso. Observa-se frequentemente jogadores que possibilitam muitos *turnovers* aos adversários por correrem a uma velocidade que não permite aplicar a sua própria capacidade de leitura e antecipação das situações de jogo. Esta é uma velocidade que não interessa à equipa. Um jogador, ao ajustar a velocidade máxima que lhe é possível ao contexto táctico, faz mais pela performance colectiva do que o puro deslocamento em corrida. Este é mais um exemplo do equívoco que representa separar o que não pode ser separado

Por outro lado, o GPAI também só discrimina jogadores entre si não os situando num nível de desempenho. Os índices de desempenho apenas permitem comparações entre jogadores, apenas permitem afirmar que alguém tem um melhor ou pior desempenho

que outro. Falta uma referência de tipo «criterial» que permita, no jogo observado, situar o jogador num determinado nível de desempenho.

Os níveis de jogo, a partir do momento que sejam representativos de graus de competência prática de domínio dos conteúdos, podem constituir-se como metas de aprendizagem e assim, serem referenciais seguros para a avaliação.

Mas o GPAI também consome muito tempo ao professor. Não é de espantar, avaliar num jogo, várias componentes para cada jogador, é tarefa hercúlea, mesmo para um professor muito experiente e especialista.

De acordo com Rezende e Valdês (2004), na aplicação do GPAI, a definição do número de componentes a serem avaliadas está dependente do número de jogadores que se pretende avaliar e, também, do número de observadores que há para registar os dados que se querem recolher.

Quer dizer, são necessários vários avaliadores para se poder avaliar todas as componentes de cada jogador, ou então, cada avaliador só avalia algumas componentes ou, alguns jogadores.

Um professor não é capaz de avaliar vários componentes e todos os jogadores. Para tal, seria necessário muito tempo e o professor não faria outra coisa. São necessários vários observadores para que uma recolha de dados permita que se retirem inferências sobre todos os jogadores em presença e sobre qual é o desempenho global no jogo.

Deste modo, é muito difícil ensinar e avaliar ao mesmo tempo. Com tanto trabalho a registar as quantidades de acções adequadas e inadequadas, ou a pontuá-las de um a cinco utilizando critérios de desempenho, dificilmente pode ser eficaz a intervenção do professor no processo de ensino e aprendizagem do jogo.

#### **1.7.2. O Team Sport Assessment Procedure**

De acordo com Gréhaigne et al. (2005) o TSAP é uma estratégia de avaliação baseada numa abordagem holística.

O TSAP baseia-se na frequência dos vários acontecimentos ocorridos durante o jogo que reflectem os aspectos técnicos e tácticos, estando a avaliação assente em informação que quantifica somente o desempenho global ofensivo individual, em jogos de invasão e de rede (Tallir et al., 2004).

Avalia principalmente a quantidade de vezes que os jogadores recebem a bola dos companheiros e a conquistam aos seus adversários. Mas também, o que fazem depois com ela, tal como, perdê-la para o adversário, manter a sua posse efectuando um passe ofensivo com sucesso ou um passe neutro que não põe em causa os adversários e rematar à baliza (Harvey, 2006).

Gréhaigne et al. (1997) consideram que pode controlar-se a participação dos jogadores mesmo que não obtenham desempenhos elevados. O TSAP pode ser utilizado a nível didáctico e na avaliação formativa e sumativa. Os dados recolhidos também lhe conferem uma interessante utilidade na investigação.

Este instrumento pode ser utilizado, nas turmas de Educação Física, numa avaliação entre pares que pode ser considerada fiável, e que resulta do envolvimento dos alunos numa activa participação na recolha e interpretação dos dados (Gréhaigne & Godbout, 1998).

Gréhaigne et al. (2005) e Gréhaigne e Godbout (1998) consideram que há três índices de equipa que devem ser combinados: número de posses de bola, numero de remates, numero de golos. Dão o exemplo do significado que têm 20 posses de bola, com 4 remates e 1 golo, reveladores de um grande número de percas de bola. Já um número semelhante de posses de bola e de remates é revelador de progresso.

Os índices de cada jogador correspondem a: bolas jogadas [BJ], bolas conquistadas [BC], bolas recebidas [BR] e bolas perdidas [BP].

No TSAP há seis variáveis de observação diferentes que reflectem o desempenho global ofensivo em jogos de invasão.

Ganho de posse de bola:

Bolas conquistadas – interceptadas, roubadas, recuperadas após tentativa de finalização.

Bolas recebidas – recebe dum colega sem perdê-la logo de seguida.

Escoamento da bola:

Bolas perdidas – quando perde o controle dela sem ter finalizado.

Bolas neutras – passe rotineiro para um colega que não põe pressão na outra equipa.

Passe – que contribui para levar a bola em direcção à baliza adversária.

Remate com sucesso – quando é golo ou se mantém a posse da bola.

Resultantes deste procedimento de avaliação autêntico baseado na observação das acções dos jogadores durante os jogos produzem-se dois índices de desempenho: o «índice de eficiência» [IE] e o «volume de jogo» [VJ] que se constituem numa pontuação total de desempenho no jogo.

O «Volume de jogo» é definido como o número de bolas jogadas por um jogador ou equipa. Parte-se do princípio que num jogo reduzido, ter muito contacto com a bola é revelador de envolvimento no jogo e participação no desempenho da equipa. Receber ou conquistar frequentemente a bola não é considerada uma coincidência.

O volume de jogo: = número total de bolas jogadas (BJ) = bolas recebidas (BR) + bolas conquistadas (BC).

Índice de volume de jogo (VJ) = BC+BR

Índice de eficiência (IE) = VJ / 10+BP

Pontuação de desempenho = (VJ/2) + (IEx10)

Numa segunda versão do TSAP o índice de eficiência sofreu modificações (Gréhaigne et al., 2005:

– Índice de eficiência = passes + remates com sucesso / 10 + BP

Mais recentemente, Nadeau, Godbout e Richard (2008) adaptaram o TSAP ao Hóquei no Gelo. Juntaram às seis categorias originais de observação, quatro novas categorias relacionadas com o Hóquei no Gelo. Foram gravados em vídeo 12 jogos dum campeonato regional do escalão etário dos 11/12 anos, envolvendo 103 jogadores. A análise em vídeo foi feita por 3 observadores treinados. Os dados foram usados para determinar por jogador o volume de jogo por minuto e o índice de eficácia, sendo os dois aspectos combinados para obter um resultado que determina a pontuação de desempenho do TSAP. Utilizaram também, medidas adicionais de avaliação e estatísticas do torneio.

Segundo os autores, os resultados fornecem evidência da validade do TSAP. A média dos resultados foi substancialmente mais alta para jogadores avaliados como melhores e para os que acumularam mais do que um ponto por torneio.

Na conclusão do estudo, afirma-se que usando esta adaptação do TSAP pode compreender-se melhor a participação ofensiva de cada jogador. No entanto, atendendo à baixa estabilidade do resultado de desempenho do TSAP, os dados recolhidos na observação devem utilizar-se mais para avaliação formativa do que para a avaliação sumativa.

#### *A crítica ao TSAP*

O facto de se considerarem índices de equipa para além dos índices de cada jogador, permite ao TSAP um menor afastamento em relação ao todo, ou seja, à circunstância total que o jogo constitui, e representa uma visão mais holística da avaliação. De facto, a equipa constitui uma unidade maior e, por isso, pode representar um ponto de partida para melhor compreender a unidade menor que é o jogador. No entanto, o jogo não é só jogado por uma equipa, as duas equipas em presença criam um produto global que pode ser observado, e isso, não é considerado no TSAP.

Também a natureza dos indicadores quantitativos considerados no TSAP está mais perto duma visão holística da avaliação. Num jogo reduzido, consideramos ajustado partir do princípio que ter muito contacto com a bola é revelador de envolvimento no jogo e de participação no desempenho da equipa. De facto, receber ou conquistar frequentemente a

bola não é uma coincidência. A questão que emerge necessariamente é a mesma: porque é que não se procuram, antes, indicadores no jogo total, em vez de os procurar no jogador ou na equipa?

Percebe-se a dificuldade. Como encarar esta possibilidade quando não existem níveis de jogo que possam ser referenciais para a avaliação? No TSAP, tal como no GPAI, como não se encara a possibilidade de partir do todo mais global que é o jogo para a parte que é o jogador, utilizam-se índices globais que não traduzem um nível de desempenho. O instrumento fica-se por uma discriminação do nível dos jogadores uns em relação aos outros, não situando cada jogador num nível de desempenho.

Afirma-se com alguma frequência que o TSAP não avalia todo o desempenho no jogo, característica que é inerente a um bom instrumento de avaliação. Esta afirmação parece-nos claramente desajustada. Não é válido o argumento de que o TSAP se fica, apenas, pela avaliação do desempenho ofensivo com bola.

Tais considerações só acontecem porque existe uma visão fragmentada do jogo. Só pelo facto dos indicadores quantitativos dizerem respeito a acções na presença da bola, isso não lhes confere um carácter exclusivamente ofensivo.

Vejamos: conquistar a bola é necessariamente o resultado de acções e movimentos defensivos que permitem, que essa conquista se verifique. Sendo assim, a intercepção tem um carácter ofensivo, é um elemento ofensivo, mas resulta da acção da defesa não do ataque. Outro exemplo: receber a bola faz parte do «jogo sem bola» ou é uma acção «com bola»? Se um jogador recebe a bola, isso resulta das suas acções e movimentos sem bola de apoio ao portador. De facto, receber a bola corresponde à criação de linhas de passe que as desmarcações proporcionam e, por isso, faz parte também da acção do jogador sem bola que é escolhido pelo passador.

Se, em primeira mão, se avaliar o todo que é o jogo, os indicadores que se procuram encontrar para ajudar neste propósito, não estarão tão sujeitos a estes equívocos porque as partes não serão vistas isoladamente.

Mas o TSAP também consome muito tempo e é pouco amigável para o professor. É muito difícil conciliar a pretensão de avaliar e ensinar ao mesmo tempo. Registar as frequências dos acontecimentos em cada um dos seis indicadores quantitativos, não deixa espaço para que todos os jogadores sejam avaliados por um mesmo observador, nem deixa espaço para o professor ter, simultaneamente, atenção na observação e registo e intervenção eficaz no processo de ensino e aprendizagem.

Considerar a possibilidade de avaliar a produção colectiva global que é o jogo, e não o jogador, reduziria drasticamente o número de indicadores a observar. Uma solução possível era ajustar os indicadores de equipa, número de posses de bola, número de



finalizações e número de golos, não exclusivamente à equipa, mas ao jogo produzido pelas duas equipas. Se os golos não forem protagonizados, ou seja, da exclusiva responsabilidade de um ou dois jogadores, poder-se-ia utilizar esses indicadores quantitativos. Bastava tentar perceber quais as frequências que revelam um determinado nível de jogo.

Por outro lado, é meritório o esforço de fazer participar os alunos na avaliação. É positiva, a perspectiva de que a utilização do TSAP pode trazer benefícios na autoavaliação e autorregulação do seu trabalho e do seu desempenho. No entanto, é duvidoso que esses dados possam ser utilizados para uma função diferente da avaliação formativa designadamente na avaliação sumativa.

Põem-se problemas de fiabilidade que não podem ser desprezados. Mesmo que os alunos sejam treinados para tal, pode não ser fácil estar claramente por dentro do constructo de cada categoria a ser avaliada ou conseguir observar o comportamento que se quer avaliar. Não sendo prudente utilizar esses dados na avaliação sumativa, já utilizá-los na avaliação formativa pode ter um grande valor para que os alunos compreendam melhor a dinâmica de jogo, as situações e as opções tácticas, no sentido de perceberem as suas possibilidades e limitações e o que devem fazer para melhorar o seu desempenho.

### **1.7.3. Outros**

Tallir et al. (2004) tentaram conceber um instrumento de avaliação do jogo, válido e fiável, muito semelhante ao GPAI. Para isso, realizaram um estudo preliminar utilizando um método baseado em vídeo.

Este instrumento difere do GPAI porque ao definir os diferentes componentes da execução ou da tomada de decisão, não os coloca numa só categoria. Também difere do TSAP porque avalia, não só, todas acções observáveis com bola, mas também, as sem bola. Em jogos reduzidos no Andebol e Futebol, os autores avaliaram constructos como passe, conduzir a bola, drible, tentativa de finalização, acção de corte, criação de espaço, defender.

Tallir et al. concluíram que o instrumento era extenso, sendo algumas das suas definições dos constructos muito gerais. Os resultados também não discriminam diferentes níveis de desenvolvimento dos jogadores. Não só há alunos que podem não ter a capacidade de executar determinadas técnicas, como também, as decisões que se tomam não estão ao mesmo nível, umas são melhores que outras. Assim, concluem pela necessidade de ajustar o instrumento a diferentes níveis de jogo.

#### *A crítica*

A crítica efectuada ao GPAI é aqui ampliada. De facto, os autores no esforço de melhorar e aprofundar a avaliação autêntica do jogo, separaram ainda mais as partes

criando novas categorias e um instrumento extenso. Mas também aumenta a dificuldade de separar os níveis de desenvolvimento dos jogadores, porque não se parte do grupo para o indivíduo, do todo para a parte, do jogo para o jogador, sem o enquadramento de níveis de desempenho que sejam referência para a avaliação do jogo e dos jogadores.

## **1.8. Como produzir classificações válidas e fiáveis do domínio dos JDC**

### **1.8.1 Situações de avaliação baseadas no jogo – um contexto autêntico para a avaliação**

O jogo formal é o contexto caracterizado pelo realismo e autenticidade das situações em que se expressam as habilidades, conhecimentos e capacidades.

Só nas situações baseadas no jogo há oposição, com duas equipas a querer exactamente o mesmo (Lames & McGarry, 2007). Só nas situações baseadas no jogo se dá o melhor pelo grupo (Gréhaigne et al., 2005), havendo dimensões sociais que determinam o nível da cooperação que permite enfrentar os opositores. Só nas situações baseadas no jogo cada jogador está dependente daquilo que os outros jogadores fazem (Gréhaigne et al., 2005), ou seja, do desempenho colectivo de colegas e adversários. Só nas situações baseadas no jogo existe um ambiente incerto (Bom, 1980; Lames & McGarry, 2007) e imprevisível (Mesquita, 2000; Garganta, 1998) que requer adaptação permanente. Só nas situações baseadas no jogo há sempre escolhas para fazer e a necessidade de colocar a habilidade ao serviço dessas escolhas. Como refere Garganta (2009), se fizermos as perguntas certas no momento certo o jogo responde a todas as nossas questões.

Numa situação de jogo, qualquer que seja a forma que assuma, a dinâmica é tal que requer eficiência na execução das habilidades bem como conhecimento táctico (Oslin et al., 1998), sendo a tomada de decisão permanente. Descontextualizar os dados de utilização de uma habilidade representa uma incapacidade de apreender a expressão total da própria actividade que é o jogo. Só a interpretação do uso das habilidades em situação, garante que se avalie a qualidade do seu contributo para a resolução dos vários e complexos problemas que ocorrem no jogo.

A actividade referente do jogo e da aprendizagem e do valor dos JDC não são exercícios de habilidades. Podem essas habilidades ser consideradas competências, pode haver concursos de lançamentos, de drible, de remates, etc. Mas a referência e o valor da actividade está nesse jogo total, nessa circunstância total. Se fraccionarmos o jogo nas suas componentes, em tarefas discretas, perdemos a dinâmica do jogo. Nazario (2005) refere, mesmo que avaliar as habilidades no jogo é mais rápido do que avaliar isoladamente as habilidades de cada aluno, sublinhando que o jogo fornece a evidência que o aluno não só sabe como fazer a habilidade, mas também sabe usá-la.

Por jogo não se entenda necessariamente desporto. O termo jogo refere-se à actividade culturalmente relevante, jogo desportivo formal, e também, como referem Rink, French e Tjeerdsma (1996), às formas simplificadas do jogo caracterizadas por poucos jogadores, regras menos complexas e equipamento modificado.

Nos PNEF, os objectivos gerais de ciclo expressam a relevância que o jogo tem, constituindo-se numa referência para a aprendizagem e a avaliação. Tomando como exemplo o 3º ciclo de escolaridade, verificamos que este objectivo geral reflecte, inequivocamente, a importância atrás referida:

“Cooperar com os companheiros para o alcance do objectivo dos Jogos Desportivos Colectivos, realizando com oportunidade e correcção as acções técnico – tácticas elementares em todas as funções, conforme a oposição em cada fase do jogo, aplicando as regras, não só como jogador mas também como árbitro” (Bom et al., 1989, p. 13).

Também os níveis de sistematização dos conteúdos dos PNEF, ou seja, as competências específicas de domínio das matérias, elegem o jogo como actividade central e decisiva. É nessa referência ou matriz que se baseia a forma das situações de prática propícias ao aperfeiçoamento e avaliação dos alunos. As condições de realização dos objectivos específicos dos programas são inspiradoras das situações de avaliação porque determinam o contexto em que as competências de um determinado nível devem ocorrer. Por exemplo, jogo de Andebol 5X5 num campo reduzido com 38x18m com bola afável nº 0, ou jogo de Voleibol 4X4 num campo reduzido 12X6 com rede a 2,15m (Jacinto et al., 2001).

O jogo tornou-se na referência principal para a aprendizagem dos JDC (Barreto et al., 1980; Graça & Mesquita, 2007). Na perspectiva da «aprendizagem cooperativa» em EF, Dyson et al. (2004) explicam que os jogos fornecem um autêntico contexto para a avaliação e argumentam que avaliar os alunos durante os jogos é a maneira mais significativa de eles desenvolverem competências como jogadores.

Harvey (2006) enfatiza, mesmo, a relação estreita entre a aprendizagem e a avaliação, afirmando que, em coerência, ensinar através do jogo trás como consequência ter que avaliar os desempenhos no jogo.

Tomando como referência as características específicas dos JDC e o significado do que representa um bom desempenho no jogo, consideramos que as situações de avaliação autêntica nos JDC, ou seja, as situações de avaliação em contexto e que estão próximas da actividade cultural referente que é o jogo, devem ter três qualidades principais:

1º - Serem significativas para os alunos, representando uma visão aproximada do que eles observam, quer na televisão., quer em situações ao vivo. A maior parte dos alunos não começa a jogar sem uma referência anterior do que é a prática formal do jogo. As modificações de regras, número de jogadores e equipamento não devem descaracterizar essa visão global do jogo.

Segundo Kirk e McPhail (2002), é importante que esteja representada toda a versão do jogo que os alunos vêm fora da escola. Também Nazario (2005) é um autor, entre outros, representativo da ideia que na avaliação autêntica a finalidade é transferir os ganhos proporcionados pela experiência de aprendizagem para a competência de participar no jogo. Deste modo, a autora preconiza tarefas significativas que representem o desempenho no jogo, tarefas que estabeleçam ligação entre as experiências do mundo real e a escola.

2º - Permitirem avaliar todos os aspectos do desempenho no jogo. Para se perceber qual é a qualidade do desempenho no jogo são importantes, quer os aspectos ofensivos, quer os aspectos defensivos. Mas, também o jogo sem bola deve ser contemplado, e não só aquele que se observa na presença da bola.

Se o objectivo de ensinar os jogos for desenvolver o desempenho no jogo, então a avaliação tem de tomar em consideração todos os aspectos do desempenho e deve ser feita em contexto de jogo (Gréhaigne et al., 2005; Siedentop & Tannehill, 2000). Como referem Oslin et al. (1998), embora a eficácia da habilidade seja importante para a performance global no jogo ela é apenas uma das suas facetas. Decisões e movimentos feitos por jogadores que não estão de posse da bola e que servem para apoiar colegas de equipa são elementos críticos de desempenho no jogo e essenciais para o sucesso da equipa. Os autores defendem que nos jogos desportivos colectivos uma considerável porção do jogo decorre sem bola e, por vezes, longe da bola e argumentam que a avaliação da performance em jogo não deve, apenas, incluir as habilidades e decisões com bola, mas também, movimentos sem bola, para se poder contabilizar a porção do jogo que ocorre longe da bola.

3º - Terem um sistema de avaliação com todas as qualidades para avaliar no jogo, nomeadamente: permitir inferências válidas sobre o desempenho colectivo e individual; avaliar todos os alunos em presença com economia de recursos, não consumindo muito tempo a professores e alunos, utilizar situações de prova que possam inspirar a avaliação formativa, permitindo que cada professor avalie e ensine ao mesmo tempo.

Na verdade, como explica Harvey (2006), ensinar através do jogo implica avaliar no jogo, com um sistema que tenha todas as qualidades necessárias a permitir, isso mesmo, avaliar no jogo.

### **1.8.2. Os grandes problemas de avaliação nos jogos**

O contexto autêntico onde se avaliam todas as habilidades, conhecimentos e capacidades para jogar é o próprio jogo. As suas características tornam a tarefa de avaliar muito difícil e, se não forem devidamente tomadas em consideração, podem comprometer a fiabilidade das avaliações. Assim passemos a analisar essas características e discutir algumas das suas implicações.

### *Abertura, incerteza e imprevisibilidade*

Abertura, incerteza e imprevisibilidade são resultantes da enorme complexidade que caracteriza um jogo desportivo colectivo. O jogo é um sistema complexo, contendo muitas variáveis interactivas (Gréhaigne et al., 2005). Nesse sistema complexo, que é composto de subsistemas com interacções dinâmicas entre eles, cada jogador tem várias alternativas de actuação, (Lames & McGarry, 2007).

Os jogos de invasão são muito imprevisíveis, pela variedade e a quantidade de relações entre colegas de equipa e adversários, o número de variáveis e a complexidade das acções (Mesquita, 2000), não se conseguindo prever com antecedência o que vai acontecer, nem a ordem cronológica desses acontecimentos, que surgem em catadupa, o que torna muito incerta a evolução sequencial do jogo (Garganta, 1998). Daí a importância de produzir incerteza nos adversários e reduzi-la para os colegas de equipa (Gréhaigne et al., 2005).

Neste primado da incerteza, a regra é contar com o inesperado (Bom, 1980) ou, seguindo a expressão de Lames e McGarry (2007), com a sorte, o caos, o clímax.

Com as condições permanentemente a mudar, as habilidades técnicas designam-se de abertas porque estão sujeitas a variações de ritmo, intensidade e amplitude gestual e, dependentes dos problemas colocados pelo jogo (Mesquita, 2000). “Enquanto habilidades abertas, elas são então reguladas pelos constrangimentos dos factores exteriores (posição e movimentos dos colegas e adversários, colocação no terreno de jogo, distância do alvo a atacar ou defender, entre outros)” (Graça, 1998, p. 28).

Nos JDC onde a invasão está ausente a previsibilidade é maior. Para Mesquita (2000) é o que ocorre, por exemplo no Voleibol, onde a utilização da técnica está mais referenciada aos modelos de execução.

Em suma: o acerto nas decisões está permanentemente a ser posto em causa, pela imprevisibilidade e complexidade da movimentação dos colegas de equipa e, principalmente, de adversários.

### *Desempenho colectivo*

Como o desempenho é colectivo, isto representa uma das maiores dificuldades de avaliação nos JDC. Cada jogador está muito dependente do desempenho de todos os outros intervenientes. Gréhaigne et al. (2005) consideram que a avaliação de cada jogador nos JDC é muito difícil, porque nas relações de oposição e cooperação, interagem inúmeros elementos, podendo a harmonia de força variar muito. Para os autores, os elementos de uma equipa são interdependentes, defendendo que para haver coerência cada jogador deve ser simultaneamente avaliado em conjunto com a equipa. Também Nevett, Rovegno, Babiarz e McCaughtry (2001), sublinham que as decisões dos jogadores num jogo não

podem ser analisadas isoladamente, mas sim tomando em consideração o contexto das acções dos colegas de equipa e dos oponentes.

Conclui-se que conseguir determinar aquilo que é crítico para avaliar pode não ser suficiente para ter sucesso. Muitas vezes, os contextos em que se pretende avaliar não proporcionam essa oportunidade. Alguns comportamentos podem não ser observados pela qualidade superior das oposições ou pelas diferenças significativas entre colegas de equipa.

#### *Irrepetibilidade*

O jogo é sempre o mesmo, mas é irrepetível. É o mesmo jogo em termos de estrutura, de modelo geral, mas é irrepetível. Para Gréhaigne et al. (2005) nunca há duas situações de jogo que sejam, absolutamente, idênticas.

Cada vez que os jogadores são colocados em situação de jogo produzem, na abertura e incerteza, um novo jogo. Inventam com criatividade prática um outro jogo, que é, também, sempre o mesmo. De acordo com Lames & McGarry (2007), as interacções dinâmicas que caracterizam os jogos desportivos são características chave do desempenho desportivo. Essas interacções são apertadas e mudam durante o tempo de jogo, porque que as equipas têm objectivos comuns, mas que são mutuamente exclusivos. Para os autores, o comportamento de cada jogador não é expressão de propriedades estáveis, quer sejam habilidades técnicas ou capacidades motoras, como é o caso da endurance nas corridas de longa duração ou da força no halterofilismo. Concluem que os desempenhos desportivos estão dependentes do contexto, situação e oponente, e dependentes do tempo e assim não repetíveis, sendo por isso, não fiáveis.

Conclui-se que o facto do jogo nunca se repetir complica tudo. As situações de jogo são sempre diferentes mesmo que apresentem semelhanças e, as acções a observar, podem não ocorrer com a frequência necessária ou, nas situações que mais nitidamente as revelam

#### *Instabilidade dos indicadores de desempenho*

Fazendo o ponto da discussão teórica, Garganta (2009) refere que um indicador de desempenho é a selecção, ou combinação, de variáveis de acção que têm como objectivo definir alguns ou todos os aspectos de um desempenho e que, é normal, usarem-se frequências dos comportamentos ocorridos para relatar indicadores de desempenho. Sublinha que há dificuldade em interpretar dados numéricos e visuais, mas também aspectos situacionais como processos físicos e cognitivos, a qualidade do oponente e o nível de preparação.

De facto, a estatística descritiva omite as importantes interacções dinâmicas que se estabelecem, não contém informação do contexto sequencial do jogo, nem revela a situação de contexto duma acção específica (Lames & McGarry, 2007).

Por outro lado, continuando a fazer o ponto da discussão teórica, Garganta (2009) defende que os comportamentos têm significado se quebrarem equilíbrios ou revelarem estabilidade nas várias acções, porque os comportamentos de jogadores e equipas são gerados a partir da tensão entre regularidades e a produção de novidade. Para o autor, alguns estudos tentaram analisar o desempenho em termos de identificação de características e variáveis que caracterizam os padrões de jogo, no entanto estas análises restringem-se às situações que conduzem à finalização e permitem somente uma compreensão muito restrita das dinâmicas de jogo e dos desempenhos das equipas

Seguimos com Lames e McGarry (2007), concluindo que o problema maior é que dada a instabilidade das características de desempenho do jogador no contexto, ao tomarmos tais características como medidas ou indicadores, obtemos classificações que não são fiáveis.

### **1.8.3. Como produzir classificações válidas e fiáveis**

O jogo é válido, mas como produzir classificações válidas quando a entidade a ser avaliada é instável, ou seja, o desempenho é instável e ocorre num contexto também ele instável?

E como fazê-lo com economia de meios, ou seja, sem consumir demasiado tempo, nem envolver directamente demasiadas pessoas, permitindo que se possa avaliar, observar e ensinar ao mesmo tempo?

Fazendo o ponto da discussão teórica, Nazario (2005) refere que a avaliação alternativa pode consumir muito tempo, e até, controlar a vida do professor. Argumenta que uma prova de duração longa tem grandes probabilidades de não ser utilizada pelos professores. Conclui que este factor pode determinar que os professores não apliquem novas formas de avaliar.

A solução adequada tem, então, de ser simples, porque normalmente só as soluções simples são económicas e eficazes. Tem de ser uma solução que, em situações de avaliação autêntica, em matérias especialmente difíceis de avaliar como são os jogos desportivos colectivos, permita que se façam apreciações válidas sobre o desempenho dos alunos, sem ocupar muito tempo, nem utilizar procedimentos de avaliação e registo trabalhosos.

A resposta é avaliar partindo do grupo para o indivíduo, do todo para a parte, da síntese para a análise e do jogo para o jogador.

#### **1.8.3.1. Do grupo para o indivíduo**

“Although sociologists and social psychologists who have studied groups have always claimed that a group is greater than the sum of its parts, there has been a tendency in assessment to regard the potential for performance in a group as the sum total of the amount of information, skills, and abilities that individuals bring to the group” (Cohen, Lotan, Abram, Scarloss & Schultz, 2002, p.1046).

Para os autores, há uma aprendizagem de grupo que resulta da interacção dos seus membros e que vai para além da capacidade de cada um. Referem, também, que nos grupos a resolução de problemas e a construção do conhecimento é conseguida se os seus membros trocarem ideias de forma criativa, não se tratando, só, de dividir trabalho para cada pessoa contribuir com diferentes peças para o produto final. Segundo Scarloss (2002), para além da contribuição individual, há qualquer coisa que não se consegue exprimir por palavras pela qual o indivíduo consegue superar o que seria capaz se trabalhasse sozinho. Para o autor, se o grupo é diferente da soma das partes, uma vez formado torna-se uma entidade coerente e única.

Na perspectiva de Crahay (2002), é necessário que a expressão do sucesso colectivo tenha repercussão em cada um dos alunos. Espera-se que todos ganhem e que o ganho individual seja proporcional à qualidade do trabalho de grupo. Para Cohen et al. (2002), quer a qualidade da interacção entre os membros, ou seja, da discussão em grupo, quer a qualidade do produto de grupo, têm consequências não só na aprendizagem de grupo, mas também no desempenho individual, que é observado na melhoria da compreensão e na capacidade individual em expressar o que se aprende. São, por isso, bons indicadores do potencial de aprendizagem do grupo. Os autores argumentam que o produto deve ser avaliado porque tem um papel importante na mediação da instrução. Referem que determinadas discussões em grupo têm consequências na qualidade do produto de grupo e influenciam a natureza da aprendizagem de grupo.

Por outro lado, para estes autores, a aprendizagem de grupo está dependente da clareza do critério de avaliação, embora sublinhem que na literatura sobre avaliação, os resultados das investigações não são conclusivos sobre o facto do desempenho melhorar em função do conhecimento do critério de avaliação, essa importância vem da pesquisa em psicologia social. No entanto, King e Behnke (2005) referem que os alunos apreciam estar por dentro dos métodos de classificação, motivo pela qual devem ter a oportunidade de discuti-los e questioná-los.

Será, então, o produto de grupo, uma medida válida do desempenho?

Bracken (2002), refere que se os procedimentos de avaliação estiverem apenas baseados no produto, os melhores alunos esforçar-se-ão de forma alcançarem uma boa classificação, dando aos alunos fracos a oportunidade de obter uma classificação inflacionada. O autor defende que as classificações individuais devem representar o desempenho e a contribuição do aluno, não o produto produzido em grupo.

No entanto, na sua leitura da matéria publicada, Scarloss (2002) considera que a maioria dos autores não encara a possibilidade da avaliação realizada em sala de aula incluir algo para além do desempenho individual e refere que, quando os dados de



avaliações de grupo são utilizadas, dizem respeito à agregação de dados relativos a avaliações individuais. O autor sublinha que os produtos de grupo não são intrinsecamente empreendimentos individuais, e defende, por isso, que não se devem agregar pontuações individuais.

Num estudo que comparava o desempenho em pequenos grupos e o contexto de avaliação individual de competências na matemática, Webb (1993) concluiu que todos os alunos tiveram um desempenho de alto nível na situação de grupo. O perfil da interacção de grupo permitiu um melhor prognóstico da prestação individual do que o desempenho individual na situação de grupo.

Também Cohen et al. (2002), numa investigação realizada com 163 alunos, estudando uma unidade didáctica de Filosofia e História da Antiguidade, confirmaram três hipóteses:

1- A utilização do critério de avaliação conduz a uma interacção entre os alunos mais avaliativa e focada na tarefa, bem como a um melhor «produto de grupo», e está associada a bons resultados numa prova final escrita.

2- Mais auto-avaliação e melhor qualidade do «produto de grupo» estão associados a uma média mais alta dos resultados da avaliação escrita.

3- Quanto melhor for a qualidade da interacção de grupo e a qualidade do produto de grupo, melhor será o desempenho individual de cada membro.

Na Educação Física, Dyson et al. (2004) consideram que a melhor forma dos alunos dominarem os conteúdos dos JDC, é trabalharem, colaborativamente, em pequenos grupos compostos heterogeneamente.

Na verdade, o grupo representa um bom contexto para aprender. Só que nos jogos colectivos a equipa não é só o contexto em que se realizam as aprendizagens, a equipa faz parte da essência do jogo. Na EF as matérias de grupo só se aprendem em grupo, o aperfeiçoamento e desempenho individual estão dependentes das práticas colectivas porque o desempenho é intrinsecamente colectivo.

Se, até em outras matérias do currículo, o produto de grupo prediz o desempenho individual (Cohen et al., 2002; Scarloss, 2002; Webb, 1993) e nas matérias colectivas da EF, o desempenho é intrinsecamente colectivo, logo, em coerência com esta constatação, a avaliação nos jogos desportivos colectivos deve considerar o desempenho colectivo como ponto de partida para avaliar o desempenho individual.

#### **1.8.3.2. Do todo para a parte**

Outro aspecto importante, a considerar, é que da mesma forma que o produto colectivo não é a soma de desempenhos individuais, também a avaliação das actividades de

grupo não é a soma das partes, embora estas possam ter notação própria, como a gramática pode ter na Literatura, ou o desenho pode ter na Pintura.

Não se avaliam actividades de grupo, que são válidas, avaliando as suas parcelas e depois somando-as. As partes podem ser fiéis e objectivas mas não são válidas. Pode-se recolher informação sobre uma parcela, que se repete independentemente das situações e de quem avalia, mas isso não autoriza a que se considerem válidas as inferências que se pretendam fazer sobre o todo ou o grupo. Só o todo permite perceber as partes e a sua soma.

De acordo com Elias (1997), as partes constitutivas de um todo são interdependentes, estando organizadas e integradas no todo. Quando isoladas da unidade de que fazem parte, perdem o seu carácter específico, alterando radicalmente a sua própria estrutura ou a estrutura da unidade maior, ou seja, perdem a sua estrutura própria porque o todo ou a síntese que as une se desfaz. Para o autor, medir ou quantificar partes isoladas, só é possível e adequado quando no todo essas partes não estão interligadas, ou estão-no levemente.

Se o desempenho do grupo é a referência de validade, trata-se, então, de perceber em que condições, a sua avaliação pode ser fiel e objectiva. O jogo na sua complexidade funcional não é susceptível de ser reduzido às particularidades das partes que o constituem, nem explicado apenas através do funcionamento e comportamento destas últimas, mas tem antes de ser esclarecido com base no carácter próprio da configuração que as partes constroem em conjunto. Numa perspectiva epistemológica, Elias (1997) argumenta que apesar do estudo das partes, poder ser necessário e útil, se não for incorporado no estudo do todo pode conduzir a conclusões erradas. Afirma que não basta conhecer as particularidades das partes, para explicar as particularidades do todo mais complexo. E concluindo, defende que, em vez de se pensar nas partes para chegar ao todo, deve-se pensar no todo para chegar às partes.

Se dividirmos o jogo em partes, perde-se a essência do próprio jogo. Desfazendo a organização, dissolvendo toda a complexidade, desaparece a ordem segundo a qual as partes estavam ligadas, perdem-se as características, já que as partes só têm valor quando relacionadas e ligadas na dinâmica da circunstância total, que é o próprio jogo. Revendo a literatura, Silva (2008) põe em causa estudos no Futebol cujos processos de análise dividem em partes os aspectos que caracterizam o jogo. Para o autor, ao não manter o essencial da relação entre as partes, os investigadores afastam-se daquilo que é a essência e a especificidade do jogo.

Na história recente do ensino nos jogos colectivos, em que a tendência dos vários modelos (TGfU, SE, e outros) é que os alunos se envolvam no jogo real, o compreendam e

procurem resolver os problemas que ele lhes coloca, sendo as habilidades técnicas, ou seja, os componentes, aprendidos em função da necessidade reclamada pelo jogo. Parte-se do todo para a parte, da síntese para a análise e não o contrário. É assim, que também a avaliação deve ser, recorrendo a métodos e técnicas de observação e avaliação que sejam mais coerentes com um ensino autêntico e que não retirem significado ao todo.

#### **1.8.3.3. Da síntese para a análise, do jogo para o jogador**

Consideramos três tópicos numa ordem de precedência, a saber: níveis de jogo, critérios de avaliação e indicadores de observação

##### *a) Níveis de jogo*

No jogo há um produto colectivo. Quaisquer que sejam os intervenientes, há um produto, uma produção colectiva, que é fruto da complexa conjugação de acções entre todos, colegas de equipa e opositores, que é observável e reflecte um determinado nível de jogo.

Se estamos perante uma produção colectiva que traduz um nível, é, precisamente, nesse nível de jogo que o conjunto dos alunos em presença há-de estar. Decidido qual o nível de jogo observado, bastará destacar quem não se encontra nesse nível, porque faz parte de um outro que é diferente do observado. Adoptar e utilizar este procedimento é mais rápido e eficaz do que estar a observar e a avaliar cada aluno individualmente.

Os níveis de jogo permitem situar o jogo e o jogador no jogo. Nos PNEF os Objectivos Específicos, por ano, definem para cada matéria três níveis de especificação e domínio – Introdução [I], Elementar [E] e Avançado [A]. Ao traduzir a actividade representativa de um certo grau de competência prática de interpretação dos conteúdos dessa matéria, os níveis revelaram-se uma referência fundamental para a avaliação porque categorizam o jogo, permitindo situar a qualidade do jogo e a qualidade do jogador no jogo.

É muito sentida, a nível internacional, a falta de níveis de jogo estruturados que sirvam de referência para a avaliação nos jogos desportivos colectivos.

Dyson et al. (2004) defendem que níveis de jogo com diferente complexidade táctica podem ajudar os professores a dar resposta à necessidade de desenvolvimento do jogo dos alunos.

Apoiando-se noutros autores, Harvey (2006) refere que as unidades de trabalho devem de estar distribuídas pelo ano e em espiral, em termos dos níveis de dificuldade do jogo. Tallir et al. (2004) concluíram do seu estudo que algumas das definições dos seus constructos eram muito gerais e a natureza do próprio jogo não permitiu separar os níveis de desenvolvimento dos jogadores. Também Harvey (2004) tinha concluído, do seu estudo, que para haver uma avaliação qualitativa do jogo o desempenho deveria ser categorizado em níveis. O autor sugere que para servir de referência à avaliação poderiam ser produzidos

videoclips de crianças a jogar que fossem representativos desses níveis de jogo e, numa perspectiva mais amigável para o avaliador, desenharem-se séries de jogos modificados que fossem, também, representativos dos vários níveis de jogo.

Os níveis de especificação e domínio das matérias elaborados nos PNEF, representam a progressão do sujeito no domínio da matéria, o que podemos designar por constructo metodológico de base. Esse constructo, ou modelo, apresenta-se como um mapa para as opções didácticas dos professores, na elaboração das progressões pedagógicas que guiam as aprendizagens dos alunos no âmbito dos “Planos de Turma” (Bom, et al., 1989; Jacinto et al., 2001).

Esses níveis representam a visão de como uma matéria e o domínio dos seus conteúdos pode progredir ao longo do currículo, visão que conjuga a opinião dos autores dos programas e dos especialistas que colaboraram nesse projecto, no caso, todos professores com currículo profissional, técnico e pedagógico, nas respectivas matérias.

O conjunto do programa, incluindo esta sistematização, foi posto em discussão e análise crítica numa rede alargada de consulta, incluindo professores, escolas e especialistas, sendo o seu processo de elaboração amplamente participado. Seguidamente, de 1989 a 1993, foi desenvolvida a experiência de aplicação dos programas acompanhada pelos autores e pela rede de consultores.

Posteriormente, em 2001, os programas foram revistos num processo que voltou a contar com vários especialistas (Jacinto et al., 2001) e que também esteve em discussão pública, não se conhecendo qualquer crítica a esta sistematização de níveis.

Assim, depois de adoptados, os níveis de jogo podem ser objectivamente aplicados. Essa possibilidade é uma vantagem, que tem e deve de ser aproveitada.

Quando se quer classificar a tendência é padronizar. Mas o jogo é «não padronizado» e portanto, torna-se difícil definir uma notação ou sistema de classificação para o desempenho observado. Os níveis de jogo permitem, precisamente, classificar o jogo e situar a performance individual. Assim, quando se avalia nos JDC, pode-se situar cada aluno num dos três níveis em que se diferencia o tratamento da matéria.

#### *b) Critérios de avaliação*

Um aluno cumpre um nível do programa, no caso, «Introdução», «Elementar» ou «Avançado», quando o seu desempenho, conhecimentos e atitudes representam a totalidade do nível considerado. Assim, importa definir qual o critério a partir do qual se pode considerar que cada aluno tem o grau de competência prática de interpretação dos conteúdos, que reflecte esse nível.

Em termos gerais, os critérios de avaliação estabelecem o que o aluno deve fazer para demonstrar que domina os objectivos (Serpa, 2010). Se bem elaborados, são,

justamente, os critérios de avaliação que permitem distinguir claramente os níveis de jogo. Allen (2002), considera que um critério claro descreve o que difere na competência de um nível para outro. É de acordo com o critério que o desempenho do aluno é avaliado. Para Dejong et al. (2002), é determinante fazer acompanhar o desempenho desejado com o critério de desempenho, ou seja, definir como o bom é suficientemente bom.

“Watching game play and observing specific students` behaviors are two different things” (Gréhaigne et al., 2005, p.86). A apreciação que resulta da observação das aprendizagens dos alunos implica um critério que é a referência desse processo de avaliação, ou seja, o desempenho individual é comparado com um critério específico e não com o desempenho de outros (Hay, 2006). O autor considera ser necessário, não só, recolher informação do desempenho do aluno, mas também definir os níveis desse desempenho. No entanto, reconhece que muito pouco existe sobre padrões ou descritores que façam a distinção entre níveis.

Em conjunto com os indicadores de observação, os critérios de avaliação garantem que os jogadores e o jogo passam a ser referenciados com mais facilidade.

#### *c) Indicadores de observação*

Em avaliação, um indicador sinaliza o facto de estarmos perante o efeito esperado e é lido como o cumprimento de um critério (Serpa, 2010).

Nesta perspectiva, a utilização de indicadores de observação, que permitam observar as características do produto colectivo que representa o nível de jogo, são uma ajuda, preciosa, para determinar se os critérios de avaliação são cumpridos, ou não.

Numa avaliação autêntica, que aproveita as situações mais próximas ou idênticas à actividade referente, as abordagens de avaliação holísticas estão sempre mais perto de, com eficácia e eficiência, recolher informação que represente o desempenho visto como um todo. Para Clímaco (2005), os estudos holísticos consideram que “a realidade é complexa e que o todo, tal como o contexto, é mais importante como provocador de efeitos ou de resultados, do que cada uma das partes que o compõem, isoladamente” (p.119).

Ora, os indicadores de observação podem ser mais representativos do todo, do que as estatísticas de jogo ou, noutra opção, os juízos ou apreciações de qualidade, que correspondem a características ou partes do desempenho e estão muito dependentes do contexto de jogo e do tempo.

Os indicadores de observação, porque representam as características específicas de um determinado nível de jogo, podem ser mais consistentes na representação global do desempenho contextualizado nesse nível, em vez de se chegar ao nível de cada jogador por agregação ou justaposição de notações de desempenho isoladas e individualizadas.

## **Capítulo 2**

### **Metodologia e Plano de Pesquisa**

Conforme está referido na Introdução, para desenvolver o estudo focado nas hipóteses de trabalho era necessário garantir três condições metodológicas, a saber:

1ª- Tomar como referência empírica o modelo ou padrão de avaliação de uma escola, caracterizada por ter práticas de avaliação autêntica em EF alinhadas com os níveis e objectivos dos PNEF, com o objectivo de verificar a validade e fiabilidade dos procedimentos de avaliação. Essas práticas de avaliação autêntica necessitam estar suficientemente experimentadas e apuradas, permitindo operacionalizar os procedimentos de avaliação para que uma amostra alargada de professores os utilize nas tarefas de avaliação autêntica dos JDC. Tais tarefas, são baseadas na discriminação de níveis de jogo, em condições controladas de (i) formação no domínio das matérias e respectivos níveis programáticos e ainda (ii) de treino de observação das qualidades do jogo e do desempenho dos jogadores, segundo esses níveis.

Esta condição é organizada e explicada no ponto 2.1 deste capítulo.

2ª- Utilizar objectos de avaliação classificados em função de um «padrão de ouro», que garanta classificações válidas e fiáveis que sejam referência de notação da situação de jogo e de cada jogador, em relação à qual possam ser comparadas as classificações produzidas pelos sujeitos da amostra. Tais objectos de avaliação, que podem ser reproduzidos e avaliados repetidamente, nas mesmas circunstâncias, são os desempenhos dos jogadores/alunos em jogos de andebol na escola, que pela sua natureza irrepitível, requerem o uso de filme ou gravação em vídeo como única solução para se obter um registo permanente do desempenho e das situações de prova.

Esta segunda condição encontra-se tratada em 2.2.

3ª - Desenhar e seguir um plano de pesquisa *quasi* experimental segundo o modelo designado como «grupo de controlo não equivalente» (Tuckman, 2000; Thomas et al., 2007), com o estudo a decorrer em três etapas, tal como na experiência relatada por Noizet e Caverni (1985). Para distinguir o tratamento do grupo de controlo, dos tratamentos dos grupos experimentais, adaptou-se o protocolo de avaliação dos JDC dos professores do GEF da ESDLG (cf. Anexo 1), especificamente para a aferição de critérios de classificação.

Em 2.3 especificamos a resolução delineada para esta condição.

Tomando como referência estas três condições metodológicas passamos, de seguida, a descrever a forma como foram garantidas.

## **2.1. Provas de Avaliação Autêntica no GEF da ESDLG**

Para dar resposta à primeira condição metodológica, tratava-se de encontrar uma geração de professores que consistentemente aplicasse os programas e que tivesse uma

grande experiência de avaliação autêntica baseada neles, nomeadamente, nas suas orientações metodológicas, nas NRSEF e nos níveis Introdução, Elementar e Avançado.

O GEF da ESDLG reunia, precisamente, estas condições, se não vejamos:

- Os seus professores aplicam os PNEF;
- Faz parte do seu património e história colectiva a realização de provas de avaliação interna em EF desde 2003. São provas de avaliação autêntica com carácter de aferição e sem consequências directas na classificação dos alunos;
- Na avaliação dos JDC são utilizadas situações baseadas em contexto de jogo que se inspiram nas condições de realização dos objectivos específicos dos programas. Para além disso, os professores avaliam o grau de competência prática com que os alunos/jogadores jogam o jogo, utilizando, precisamente, os níveis de jogo, para situar os seus desempenhos.

As provas de avaliação interna em EF, especialmente, todo o processo da sua preparação e operacionalização, constituem razões válidas para se poder reconhecer que o GEF da ESDLG revela grande experiência, conhecimento e compreensão dos aspectos conceptuais, metodológicos e práticos da avaliação do desempenho dos alunos nos JDC.

Para melhor compreendermos o processo de preparação e operacionalização das provas passamos a descrever os passos mais importantes dados pelo GEF da ESDLG. Nesse processo podemos reconhecer sete passos, a saber:

1º Definem-se prioridades e identificam-se metas curriculares.

Pretende-se que os PNEF sejam uma referência para a desejável articulação entre o currículo nacional e aquilo que são as responsabilidades das «estruturas pedagógicas» da escola e a deliberação de cada professor. Supostamente, os programas garantem as condições subjectivas de equidade e harmonização das decisões dos Departamentos Curriculares e dos Conselhos de Turma em cada escola e, no conjunto das escolas.

Neste quadro, as provas de avaliação interna têm constituído uma peça importantíssima de articulação entre o currículo nacional e a responsabilidade do grupo e de cada professor, permitindo harmonizar decisões dentro da escola e, até, mediar as classificações que cada professor atribui nas suas turmas.

De acordo com o projecto de EF e com o plano plurianual do GEF da ESDLG definem-se os anos de escolaridade que são prioridade em cada ano lectivo. Também se define o momento ou momentos em que se vão efectuar as provas de avaliação interna. Por vezes, realizam-se no final do ano lectivo, nos anos terminais de cada ciclo, para obter o retrato do que os alunos foram capazes de aprender nesse ciclo de escolaridade. Outras vezes, para se poder avaliar a progressão nas aprendizagens entre dois momentos,



realizam-se no princípio do ano lectivo durante o período de avaliação inicial e no final do ano lectivo na etapa de demonstração de resultados.

Para cada ano de escolaridade torna-se, também, importante identificar o que se espera que os alunos alcancem como resultado da instrução. Durante o processo de preparação das provas os professores do agrupamento descobrem aspectos de cada matéria que devem ser enfatizados e como a instrução os pode enfatizar. Esta clarificação do currículo permite aos professores verificar que, por exemplo, no nível introdução da Ginástica Acrobática já se espera que os alunos integrem numa coreografia as figuras aprendidas, sendo necessário que isso seja contemplado no plano anual de cada turma, ou, usando outro exemplo, verificar que determinada dança faz parte dum dado nível do programa e, por isso, tem de ser ensinada nesse ano lectivo. Como refere Allen (2002), os professores tornam-se parte do processo de revisão do currículo.

Mas este, também, é o momento de colaborativamente estabelecer prioridades. De decidir quais as matérias prioritárias a avaliar e que níveis contemplar. Segundo Popham (2003), é através do desempenho evidenciado pelos alunos numa prova que se fazem inferências sobre conhecimentos, habilidades e atitudes. Para este autor, a prova deve permitir avaliar os conteúdos que têm um grande significado e que, por isso, são considerados altamente prioritários. Em contraponto com aqueles criticam a estratégia de prioridades, afirmando que o professor pode estar a dar atenção, apenas, a poucas metas curriculares, este autor argumenta que não basta, simplesmente, tentar ensinar conteúdos, mas há que ensiná-los verdadeiramente.

Trata-se, então, de em determinados domínios do currículo considerados importantes, se avaliarem as competências consideradas prioritárias. No caso da ESDLG, de acordo com o plano plurianual do GEF, significa avaliar-se as matérias nucleares dos programas e os desempenhos dos alunos em níveis de matéria considerados prioritários.

## 2º Reflete-se sobre possíveis «perversões».

Há dois fenómenos perversos associados à realização de provas que poderão surgir com grande nitidez, o do estreitamento do currículo – reducionismo curricular, e o da preparação exclusiva para a prova.

Se reconhecermos a influência que a avaliação tem nas decisões e práticas dos professores, coloca-se sempre a hipótese de avaliando apenas uma parte do currículo se poder estar a provocar o seu estreitamento. Aquilo que é essencial pode tornar-se o máximo. Referindo-se a críticas que por vezes se fazem aos exames, Fernandes (2005) sublinha, principalmente, o facto de avaliarem uma amostra muito reduzida dos domínios dos programas e terem por vezes efeitos nefastos, condicionando o ensino aos modelos e matrizes de exame e assim tornando o currículo mais pobre. Como refere Conceição (1994),

a tendência de professores e alunos, é ensinar e estudar o «que pode sair» e esquecer as restantes aprendizagens. Para Popham (2003), conteúdo que não sai no exame pode ser pura e simplesmente ignorado.

Por outro lado, preparar para o desempenho numa prova não é o mesmo que preparar para o desenvolvimento das competências que se pretende que uma prova avalie. Segundo Popham (2003), deve-se ensinar preparando para os objectivos representados nas provas e não para as provas. Para V. De Landsheere (1988), muitos professores tentam estar em sintonia com o que julgam ser as preocupações dos examinadores e dedicam às tarefas de aprendizagem um tempo proporcional a essas preocupações. Para a autora, os alunos estudam, precisamente, em consonância com essa forma de actuar dos professores.

Em função do conhecimento destes fenómenos, os professores do GEF da ESDLG discutem, e.g. a perversão que representa treinar sempre com a mesma música uma determinada dança, para assegurar que durante a situação de prova os alunos não «saíam da música» ou utilizar, disfarçadamente, «defesas sombra» no futebol, para garantir que durante a situação de prova os alunos revelam bom controlo de bola e têm tempo suficiente para tomar decisões adequadas. Portanto, discutir em colectivo as questões do reducionismo curricular e as da preparação exclusiva para a prova, identificando as situações em que aquelas mais nitidamente se revelam, significa que os professores se encontram melhor preparados para a possibilidade da sua ocorrência.

### 3º Define-se o critério de sucesso

Esta definição não é fácil, nem é ausente de polémica. Como se baseia em opiniões de professores, mesmo que sejam especialistas, resulta da sua interpretação da realidade da educação em cada escola e das possibilidades dos alunos. Para a National Association of State Boards of Education [NASBE] (1999), os critérios de sucesso são uma clara medida para se saber quando «o bom é suficientemente bom» e para avaliar o progresso. Esta associação afirma que a decisão é, simultaneamente, uma decisão subjectiva e científica, e significa o mesmo política e ideologicamente que definir os objectivos dos programas, considerando, ainda, deslocada qualquer discussão à volta da forma como esses critérios poderão ser fixados. Mais importante do que a maneira como são fixados é serem válidos.

Para V. De Landsheere (1988), supõe-se que é possível especificar objectivamente, para uma determinada tarefa, um desempenho mínimo aceitável, havendo várias expressões que podem designar esse desempenho: nota de sucesso «passing score», nota de corte «cut-of-score», limiar de sucesso «threshold», nível de exigência mínimo «standard». A autora considera que é uma ilusão pensar que isto se pode determinar com

uma precisão matemática. Assim, interroga-se sobre a possibilidade de determinar com exactidão, por exemplo, o domínio de uma língua estrangeira ou a boa cidadania.

Para esta autora, reduzir a arbitrariedade e garantir coerência nas decisões, passa por estabelecer regras e procedimentos comuns, mantendo os juízos de valor e evitando a sofisticação estatística típica de quem se contenta com esforços puramente matemáticos para fixar a nota de sucesso.

Em Portugal, onde os contextos são relativamente diferenciados, as necessidades de adaptar e adequar os programas de educação física às várias realidades foram aumentando. Tal, tem provocado uma enorme diversidade do currículo implementado.

A gestão flexível do programa, sendo uma vantagem e uma das suas características mais inovadoras, não deve ter um efeito perverso – falhas sistemáticas de continuidade e de coerência no percurso de desenvolvimento dos alunos ao longo da escolaridade.

Se não é desejável que todos os alunos tenham o mesmo currículo, é particularmente importante definir um nível de distinção que represente o sucesso, apoie a deliberação pedagógica dos professores, ajude o aluno a formar uma imagem consistente das suas possibilidades e seja referência para tomadas de decisão ao nível do sistema educativo. Foi isso que as NRSEF (Jacinto et al., 2001), surgidas na sequência da revisão curricular de 2001, tentaram garantir. Os pressupostos para a sua determinação assentaram na conjugação de vários aspectos como a “grande diversidade de condições das escolas, as diferentes capacidades dos alunos e a própria filosofia do programa quanto à sua gestão e eventual diferenciação das matérias que compõem o currículo” (p. 37).

No caso da EF, faz todo o sentido que os níveis de sucesso das Provas de Avaliação Interna devam ser deduzidos directamente das chamadas NRSEF que constam nos Programas aprovados na Revisão Curricular de 2001.

É o que fazem os professores da ESDLG que alinham o seu critério de sucesso com as NRSEF. Tal como está estabelecido nas NRSEF, na ESDLG ter sucesso na área das actividades físicas numa prova de avaliação interna no 9º ano, significa cumprir 6 níveis introdução e 1 nível elementar, ou, no 12º ano, significa cumprir 3 níveis introdução e 3 níveis elementar.

#### 4º Revê-se e reformula-se o protocolo de avaliação

Neste ponto do processo de preparação das provas os professores revêem o protocolo de avaliação, verificando se as situações de avaliação escolhidas estão alinhadas com as condições de realização dos objectivos dos programas, se são boas situações de avaliação autêntica e se permitem distinguir os níveis de desempenho em cada matéria. Assim, nos JDC são tomadas decisões como, e.g. considerar que no Voleibol a situação de

jogo de cooperação 2 + 2 com serviço é a que melhora permite distinguir o nível introdução do elementar, ou que a situação de jogo de oposição 4 X 4 é a que melhor permite distinguir o nível elementar do nível avançado.

É, também, nesta altura que se revêem ou reformulam os critérios de avaliação e, em matérias como os JDC, os indicadores de observação. Em conformidade, para que o júri possa situar cada aluno num nível do programa, preparam-se fichas de registo de avaliação que permitam anotar de uma forma simples e precisa toda a informação considerada relevante. É nessas fichas de registo que, e.g. na avaliação da Ginástica Acrobática é anotado se o aluno em cada figura cumpre total ou, parcialmente, os critérios de correcção, ou, na avaliação de um JDC é anotado o nível de jogo observado e qual o nível de jogo dos alunos que se destacam.

Há, portanto, um especial cuidado em tornar todos os procedimentos de avaliação económicos, quer ao nível dos recursos materiais necessários, quer ao nível do tempo despendido.

“Não precisamos de mais avaliação. Precisamos de melhor avaliação! O que significa que talvez possamos evitar excessos, mais ou menos tecnicistas, nas salas de aula e fora delas, com a utilização de instrumentos que reflectem uma pulverização do currículo, como certas grelhas de observação, ou certas listas de verificação, que chegam a ter várias dezenas de itens para observar e para verificar!” (Fernandes, 2005, p. 79).

Em sintonia com estas afirmações de Fernandes também consideramos que na EF estes excessos podem representar uma perda de tempo inútil e envolver recursos desnecessários. As Provas de Avaliação Interna devem ser eficientes nos custos e no tempo que se despende a aplicá-la e a classificá-la. No entanto, algum esforço colectivo deve ser desenvolvido no sentido de harmonizar processos e técnicas de avaliação. Sobretudo, um especial cuidado deve ser tomado para que todos os procedimentos de avaliação sejam tão válidos e fiáveis quanto possível.

De acordo com Popham (2003), estes cuidados não têm de representar um grande esforço. Basta garantir que um determinado conteúdo está aceitavelmente expresso no item que o representa e o método mais adequado para o fazer é utilizar o julgamento humano. Mas, a questão pode não ser assim tão fácil de resolver. Como explicam De Ketele e Gerard (2005), na abordagem baseada em competências é utilizada a mesma técnica das provas clássicas em que os juízes verificam se os itens fazem parte do universo de referência. O problema é que nesta abordagem os itens e os universos de referência são fundamentalmente diferentes. Aqui, os itens são situações problema concretas e o universo de referência é a família de situações que corresponde à competência. Os juízes têm de verificar se as situações correspondem à família de situações, que é uma questão muito difícil senão impossível de resolver porque as situações são complexas, sendo muito difícil

verificar a sua equivalência. Para fazer bem o trabalho é indispensável dispor de uma descrição suficientemente precisa da família de situações, ou seja, de uma definição de parâmetros que a caracterizam. Para Moon et al., (2005), a validade de conteúdo pode ser garantida por um painel de professores especialistas que apreciem até que ponto as provas avaliam os objectivos de aprendizagem que pretendem avaliar e que analisem criticamente potenciais enviesamentos.

Nesta perspectiva, especialmente nos primeiros anos de realização da prova, utilizaram-se professores especialistas em cada matéria, que apreciaram a pertinência de cada situação de avaliação para recolher informação válida do desempenho dos alunos.

Por outro lado, em provas desta natureza estão incluídas coreografias de ginástica e dança, jogos de equipa e jogos em pares ou singulares. Os alunos realizam tarefas abertas, contextualizadas, em que para resolver problemas complexos, individualmente ou em grupo, necessitam de realizar um esforço significativo que solicita vários tipos de desempenho. Popham (2003) considera que este tipo de avaliação, implica um tempo significativo de investimento para professores e alunos e recomenda a sua utilização, apenas, na avaliação de objectivos altamente prioritários.

Nas provas baseadas no desempenho as rubricas podem ajudar no esforço de classificação. As rubricas de avaliação devem estar claramente definidas e serem fáceis de usar. Para Doering (2000), as apreciações dos professores não são baseadas em critérios puramente subjectivos, mas em orientações de pontuação específicas, chamadas rubricas.

Uma rubrica de pontuação define o critério pelo qual o desempenho ou produto é julgado, consistindo em níveis de desempenho e numa lista de características que descrevem o desempenho de cada nível (Siedentop & Tannehill, 2000). Para Popham (2003), as rubricas para além de serem utilizadas na classificação das respostas, podem, também, orientar os alunos na aprendizagem das competências a serem avaliadas.

É de acordo com estas preocupações que desde o início do ano lectivo os professores procuram que os alunos conheçam as prioridades para esse ano, as matérias a serem avaliadas, as competências que integram cada nível de desempenho e as situações de avaliação e respectivos critérios de classificação.

#### 5º Garante-se objectividade

Para que Provas de Avaliação Interna sejam úteis é importante ter dados sobre o desempenho dos alunos em que se possa confiar. Quem tem responsabilidades, quer ao nível do Departamento e da Escola, quer ao nível da Comunidade e do Sistema Educativo precisa de confiar nesses dados para interpretá-los e tomar decisões.

Mas, os professores do GEF da ESDLG sabem que podem olhar para o mesmo desempenho e apreciá-lo de maneira muito diferente. Sabem que na aplicação dos critérios

de classificação, se estiverem isolados perante a sua realidade e as suas opções individuais de avaliação, o seu comportamento pode revelar insegurança e relativismo. Como refere V. De Landsheere (1988), quando trabalham isolados uns dos outros é comum o desacordo entre avaliadores.

Assim, tentando concertar-se na avaliação, os professores do GEF reúnem-se e estabelecem um plano de aferição de critérios com o objectivo de avaliar em comum o desempenho dos alunos em matérias caracterizadas por imprevisibilidade e abertura, como é o caso dos JDC. Em sessões de vídeo ou através de observações em aulas, onde os alunos estão colocados em situações semelhantes às situações de avaliação, os professores discutem e resolvem as suas divergências de apreciação, combinando-se para harmonizar as suas classificações. Para Moon (2005), a concordância é o grau em que dois classificadores concordam com o nível de desempenho do aluno, sendo importante saber o valor das divergências na apreciação das respostas dos alunos, para se poder trabalhar a partir desse conhecimento.

Para que haja coerência entre os avaliadores é indispensável aferir-se, em comum, os problemas concretos de aplicação dos critérios de classificação. Isso requer que os professores coordenem e validem os seus esforços de avaliação. Não só que trabalhem em conjunto para construir as situações de prova e os procedimentos de classificação, mas também, que treinem a aplicação dos critérios de classificação para garantir que todos os avaliadores os interpretam da mesma maneira.

#### 6ª Validam-se os resultados

É muito importante validar a avaliação. A responsabilidade da avaliação estar apenas centrada nos professores titulares de cada turma pode não ser entendido como credível pelos colegas e por quem quiser interpretar os dados recolhidos.

Em cada turma, dos elementos que compõem o júri, fazem parte dois ou três professores de EF da escola. O professor titular da turma não se envolve em tarefas de avaliação porque não faz parte do júri. Tem a função de organizar a turma para que o júri possa preocupar-se, apenas, com as tarefas de avaliação. Se por questões de natureza operacional isto não for possível, está previsto utilizarem-se painéis de validação com outros professores, o que ainda não aconteceu até ao momento.

Como é aconselhado por Allen (2002), quando se usa um painel de avaliadores, podem seleccionar-se, casualmente, um grupo de desempenhos, enquanto se recolhem dados de todos os alunos. Os dados recolhidos pelos professores são enviados para o painel, que usa a mesma rubrica para avaliar os mesmos desempenhos. Este autor, refere que as diferenças de classificação podem ser identificadas e corrigidas sem grande esforço

e que podem conseguir-se coeficientes de fiabilidade altos se os professores forem adequadamente preparados.

Nos primeiros anos de realização da prova, foram utilizados professores especialistas nos JDC para validar as classificações do júri. Com o objectivo de alinhar as classificações de todos os avaliadores, as divergências mais grosseiras eram discutidas e corrigidas com a ajuda dos especialistas.

Durante as provas, também, o desempenho dos alunos é gravado em vídeo. Em situações de divergência classificativa, entre elementos do júri, utiliza-se a análise em vídeo para resolver essas divergências.

A importância de validar a avaliação resulta da necessidade de garantir que o processo de recolha e gestão dos dados é de tal forma eficaz que os resultados são entendidos como credíveis para colegas e para todos os que tomam decisões.

7º Agregam-se os dados, interpretam-se e tomam-se decisões

Os dados são agregados por ano, turma, género, populações específicas, domínios/matérias do programa, etc. É essa agregação dos dados que permite que eles sejam muito úteis, nomeadamente, para:

- Catalisar discussões sobre a eficácia da implementação do projecto de EF e dos próprios PNEF;
- Questionar métodos específicos de instrução e práticas inovadoras e verificar o seu impacto nos resultados;
- Analisar diferenças de contexto, tais como, instalações, equipamento, dimensão da turma e a sua relação com as diferenças de resultados;
- Analisar as necessidades de populações específicas;
- Revelar pontos fortes e fracos nas aprendizagens, a partir dos resultados obtidos pelos alunos nos vários domínios do programa.

Na sua leitura da matéria publicada, Clímaco (2005) considera que, apesar dos dados poderem ser válidos, significativos e úteis e, até, terem tratamento estatístico sofisticado, só há avaliação, propriamente dita, quando se questiona e interpreta a informação recolhida nos processos de avaliação,

A mesma autora considera que quando se interpretam dados, principalmente no início dos processos de auto avaliação, podem surgir tensões ou conflitos e, por isso, deve evitar-se radicalismos, sendo mais importante a reflexão que conduz às conclusões que as informações em si mesmas.

Este é, então, o momento de aprofundar processos e resultados, de verificar pontos fortes e fracos, de reflectir sobre as práticas individuais e colectivas. Como refere Roldão

(1998), é a reflexividade que permite um profissional “analisar as suas práticas, à luz dos saberes que possui e como fonte de novos saberes” (p. 7).

Mas todo o processo das Provas de Avaliação Interna só estará concluído se soubermos agir com base nos resultados. É por isso que os professores em conselho de grupo tomam decisões de melhoria ou aperfeiçoamento do sistema de educação física da escola. Tomam decisões ao nível da didáctica, do planeamento e avaliação, dos projectos curriculares, da gestão e implementação dos recursos, da formação de professores.

Na verdade, são sempre os resultados obtidos pelos alunos nas provas de avaliação seguintes, que confirmam, ou não, a bondade e o acerto das decisões tomadas.

## **2. 2. Procedimentos e critérios adoptados na elaboração dos vídeos de jogo**

Reunimos com os professores mais experientes do GEF da ESDLG, ou seja, aqueles que realizam provas de avaliação interna desde que elas começaram, com o objectivo de seleccionar turmas ou grupos de alunos que pudessem ser potenciais jogadores a ser avaliados.

Observámos aulas de duas turmas do 8º ano e de uma turma do 11º ano em que a matéria de andebol tinha tratamento privilegiado.

Observámos dois treinos das equipas de andebol do Desporto Escolar, um da equipa masculina e outro da equipa feminina.

Em conjunto com os professores das turmas e com os professores responsáveis pelas equipas do Desporto Escolar, seleccionámos os alunos que participariam nos jogos a gravar em vídeo.

Situámos cada um dos alunos no nível de jogo que traduzia o seu grau de competência prática de interpretação dos conteúdos do andebol e organizámos, criteriosamente, a constituição de equipas.

Com o investigador no papel de árbitro, para garantir que o jogo não era desvirtuado pela variável arbitragem, gravámos vários jogos de 10 minutos para obter diferentes produtos de jogo. De acordo com Noizet e Caverni (1985), nos estudos experimentais sobre o comportamento de avaliação, os produtos construídos têm de corresponder a todas as categorias possíveis. Por exemplo, se os testes integrarem composições, uns devem ser ricos de ideias e correctos no estilo, outros devem ser o contrário (pobres de ideias e incorrectos no estilo); alguns devem ser fortes nas ideias e fracos no estilo, enquanto outros devem ser fracos nas ideias e fortes no estilo, etc.

Assim, procedemos ao visionamento, catalogação e escolha dos jogos (*cf.* Anexo em CD), seleccionando:



- Dois jogos de nível Introdução, os jogos 2 e 4. Nestes jogos, os jogadores, manifestamente, não cumpriam os objectivos deste nível do programa, havendo apenas a destacar um aluno que evidenciava competências deste nível de desempenho.

- Dois jogos de nível Elementar, os jogos 1 e 5. Nestes jogos, os jogadores evidenciavam competências do nível Introdução, havendo apenas a destacar um aluno que não cumpria os objectivos deste nível do programa.

- Dois jogos de «fronteira», os jogos 3 e 6. Metade dos jogadores que compunham as duas equipas cumpria o nível Introdução e, exactamente, a outra metade não cumpria o nível Introdução. Tomou-se a decisão de considerar os jogos como sendo de nível elementar, porque evidenciavam quatro características principais: havia cooperação entre os jogadores de cada equipa; o protagonismo dos melhores jogadores não era fruto de acções, marcadamente, individualistas; os jogadores de nível introdução não faziam descer claramente a qualidade do jogo porque dominavam, globalmente, a desmarcação; o jogo tinha muitos golos.

Decidimos a ordem pela qual os jogos seriam visionados para que o conjunto de jogos A (jogos 1, 2 e 3) e o conjunto de jogos B (jogos 4, 5 e 6) tivessem um nível de dificuldade semelhante e, também, para que os participantes no estudo não percepcionassem qualquer lógica na ordem de apresentação dos jogos.

Porque foi o Andebol a matéria escolhida no âmbito do presente estudo?

Três pressupostos orientaram a escolha efectuada, a saber:

1 – O JDC escolhido teria de ser uma matéria nuclear dos PNEF. Na subárea dos JDC as quatro matérias nucleares são o Andebol, o Basquetebol, o Voleibol e o Futebol.

2 – O JDC escolhido teria de ser representativo das matérias mais difíceis de avaliar, ou seja, os jogos de «invasão». Nos jogos desportivos colectivos onde a invasão está ausente a previsibilidade é maior. Para Mesquita (2000) é o que ocorre, por exemplo no Voleibol, onde a utilização da técnica está mais referenciada aos modelos de execução.

3 – Relativamente ao Andebol, a habilitação e o currículo do investigador garantiam a qualidade necessária da formação a realizar nos grupos experimentais. O investigador tem o grau de Master Coach, o curso de treinadores de mais elevado nível na Europa, e foi treinador de Selecções Nacionais e também de clubes no Campeonato Nacional de Andebol da 1ª divisão.

#### **2.2.1. Fiabilidade das classificações dos jogadores por um painel**

Como está descrito anteriormente, não foi um painel de técnicos especialistas da modalidade de andebol que constituiu o chamado «padrão de ouro». Em vez disso, considerámos o padrão de avaliação do grupo de professores de Educação Física de uma escola, suficientemente apurado e experimentado, tomando-o como referência empírica em

que nos baseamos para definir o protocolo experimental a aplicar aos sujeitos que participaram no estudo.

Os objectos de avaliação foram criteriosamente escolhidos e resultaram da experiência de elaboração de vídeos a partir de provas de avaliação autêntica de JDC utilizados para aferir a aplicação de critérios de avaliação no GEF da ESDLG. Justamente, essa escolha dos objectos de avaliação representa a classificação em termos de «padrão de ouro», a qual foi relaizada pelo investigador e pelos professores mais experientes do GEF. Foi esta classificação que serviu para comparar as avaliações feitas pelos observadores/avaliadores. De acordo com Coutinho (2011), o conceito de "acuracia" está representado na comparação de uma codificação individual com a codificação padrão considerada correcta e é a forma mais exigente de avaliar a fiabilidade quando estão em causa julgamentos pessoais na análise.

Considerando-se pertinente avaliar a fiabilidade do próprio «padrão de ouro», utilizaram-se quatro professores que mais directamente estiveram envolvidos no processo de escolha dos jogadores e dos jogos a gravar.

Em duas sessões, separadas por um intervalo de dois meses, procurou-se verificar o acordo intra-observadores e o acordo inter-observadores. Em cada um dos seis jogos que compõem os conjuntos A e B (jogos 1, 2, 3, 4, 5 e 6), os quatro professores classificaram individualmente dois jogadores sorteados, um de cada equipa, ou seja, 25% do total de jogadores em jogo.

Após observarem episódios de jogo correspondentes aos 4 primeiros minutos de cada um dos jogos, os quais têm uma duração de cerca de 8 minutos, sem que houvesse qualquer comunicação entre eles, anotaram a sua classificação numa ficha registo (*cf.* Apendíce 4) em tudo semelhante à utilizada pelos observadores/avaliadores durante a experiência. Para avaliar a fiabilidade, Tabachnick e Fidell (1989) referem que é necessário analisar um número de eventos ou episódios que no mínimo correspondam a um valor de 10% do total.

Para analisar os resultados pretendia-se utilizar a «fórmula de Bellack» referida por Van der Mars (1989). Todavia, tal não foi necessário porque os resultados revelam um total acordo (100%) intra-observadores e inter-observadores e são completamente coincidentes com a classificação inicial feita pelos próprios, aquando da escolha de jogadores e da gravação dos jogos feita anos antes. Para Fleiss (1981), resultados superiores a 75% já mostram uma fiabilidade adequada.

Estes resultados mostram que as classificações dos professores do GEF da ESDLG eram credíveis em termos de avaliação autêntica dos JDC e que o seu padrão de

avaliação pode ter constituído uma referência empírica confiável para o efeito pretendido neste estudo.

### **2.3. O plano de pesquisa**

Como vimos anteriormente, só as situações baseadas em contexto de jogo nos garantem autenticidade quer no ensino, quer na avaliação dos JDC (Gréhaigine et al., 2005; Harvey, 2006, Barreto et al., 1980). São, portanto, as situações de jogo, as principais situações de prova que se devem utilizar para avaliar as produções autênticas dos jogadores/alunos.

Essas produções autênticas são, simultaneamente, criativas e colectivas, e expressam-se num ambiente caracterizado pela incerteza e imprevisibilidade (Bom, 1980; Garganta, 1998) onde o desempenho, que é intrinsecamente colectivo, é avaliado em condições instáveis e que não se repetem (Lames & McGarry, 2007).

Ora, esta natureza das situações de jogo torna as provas de avaliação muito abertas, e por isso, muito difíceis de avaliar. Como os dados são recolhidos através do método da observação, torna-se importante que eles sejam independentes do observador, o que remete para a necessidade de garantir objectividade (Lames & McGarry, 2007). Por um lado, o jogo é válido, ou seja, a sua validade está à partida resolvida, por outro lado, coloca problemas de fiabilidade que necessitam ser atenuados.

Muitas das divergências de avaliação acontecem por problemas complexos de resolver que estão centrados no observador. Por vezes, os observadores podem revelar pouco conhecimento da matéria, mau domínio dos critérios de avaliação, pouco treino de observação, ou falta de convicção de que a avaliação é, ela própria, um elemento estruturante do processo de ensino e aprendizagem.

Como explicam Noizet e Caverni (1985), nas divergências de avaliação intervêm efeitos que são sistemáticos, o que significa que essas divergências não são todas aleatórias e, põem como hipótese, que as divergências se explicam sobretudo por variáveis de situação, embora também intervenham variáveis da personalidade. Para estes autores, essas variáveis são responsáveis pelo facto de dois avaliadores não chegarem ao mesmo resultado quando avaliam o mesmo produto, sendo, por isso, importante na análise experimental identificar as variáveis de situação e tentar perceber qual é a sua importância, já que o objectivo é chegar a uma explicação causal da avaliação.

Thomas et al. (2007) sublinham que é, precisamente, com a pesquisa experimental que se manipulam tratamentos tentando estabelecer relações de causa efeito. Referem que por um lado, a causa tem de preceder o efeito no tempo e, por outro lado, a causa e o efeito têm de estar correlacionados entre si e essa correlação não pode ser explicada por outra

variável. Para Sousa (2005), o desenho experimental é necessário nessas relações de causa e efeito quando se quer avaliar novos métodos, técnicas, procedimentos, ou programas e efeitos nos sujeitos, como a aprendizagem, ou a compreensão.

Esta é, precisamente, a estrutura da nossa investigação. Primeiro, estudar de que forma um episódio de instrução sobre a classificação dos alunos nos jogos desportivos colectivos, que tem como referência os níveis do programa, pode influenciar a capacidade de avaliação dos professores para produzir classificações válidas e fiáveis; segundo, estudar de que forma uma técnica de avaliação, que procede do todo para a parte, do grupo para o indivíduo e do jogo para o jogador, contribui para melhorar essa capacidade de avaliação.

Por outro lado, Noizet e Caverni (1985) são autores, entre outros, representativos da ideia que o professor desenvolve um comportamento de avaliação quando aplica uma escala de valorização a produtos ou registos. De acordo com os autores, o comportamento de avaliação pode ser integrado numa categoria mais geral de comportamentos que são os comportamentos perceptivos e cognitivos e a avaliação deve ser entendida como uma actividade de comparação, em que se associa o produto a avaliar com um “modelo de referência inscrito nas estruturas cognitivas do avaliador” (p. 70). Nesta lógica, afirmam que, quando se verificam diferentes avaliações sobre a mesma produção, isso é consequência da existência de diferentes modelos de referência e/ou erros de classificação.

Para Noizet e Caverni (1985), a capacidade de avaliar também se aprende, sendo assim necessário estudar a relação dos comportamentos de avaliação com as tarefas de avaliação para que, em relação a esta temática, ideias pertinentes possam ser adiantadas.

Tomando como referência este quadro conceptual, estes autores propõem o método experimental para recolha de dados do comportamento de avaliação. O princípio do método consiste em mandar avaliar o mesmo conjunto de produtos a vários avaliadores diferentes e colocá-los, também, em condições diferentes, ou seja, não dar aos avaliadores a mesmas informações sobre os autores dos produtos. Explicam que é a comparação entre dois grupos de avaliadores que permite pôr em evidência a influência de determinado tipo de informações sobre o comportamento de avaliação, consistindo, então, o processo, “em introduzir um factor experimental” como ter esta ou aquela informação, haver ou não concerto entre os avaliadores, ser professor inexperiente ou veterano, “e em investigar se esse factor experimental provoca, ao nível da avaliação, efeitos sistemáticos” (p. 75). Afirmam que o método experimental permite, precisamente, “distinguir variações que são simplesmente aleatórias das variações que são variações sistemáticas e que, por esse facto, esclarecem a natureza do processo subjacente” (p. 76).

Para estudar o nosso problema central, na perspectiva dos resultados terem significado no contexto de referência, a prática profissional dos professores, particularmente, dos professores de EF, utilizámos uma pesquisa aplicada, com ambientes reais de formação, com quantidade suficiente de características dessa realidade para permitir generalizar as possíveis descobertas e, assim, garantir a validade ecológica do estudo<sup>17</sup> (Thomas et al., 2007). Como referem Noizet e Caverni (1985), cuidar que as situações experimentais representem a realidade, ou seja, que os resultados sejam válidos, é um problema epistemológico de qualquer investigação científica.

Assim, tomámos como referência a extracção e a síntese que Noizet e Caverni (1985) fazem da experiência realizada conjuntamente com Bonniol, em 1972, em que verificaram se uma concertação entre avaliadores tinha como efeito reduzir as suas divergências de classificação. A experiência foi realizada com 40 professores que corrigiam testes, que eram composições de Francês, construídos para que todo o seu conteúdo fosse controlado do ponto de vista do plano, do estilo e da ortografia. A experiência decorria em três etapas: na 1ª etapa, todos os avaliadores classificavam individualmente os mesmos exercícios; na 2ª etapa, num processo de concertação entre avaliadores, os professores do grupo experimental estabeleceram uma grelha de correcção comum e corrigiram um certo número de exercícios, confrontando as diferentes maneiras de corrigir; na 3ª etapa, os mesmos avaliadores voltavam a classificar individualmente trabalhos que recebiam para esse efeito.

Os resultados da 1ª etapa representavam as divergências iniciais entre avaliadores e os resultados da 3ª etapa as divergências finais. A diferença entre a importância inicial das divergências e a importância final traduzia o contributo da 2ª etapa para atenuar essas divergências. Foi, assim, possível estudar se a concertação entre professores os modificou ao ponto de reduzir as suas divergências quando de novo fossem trabalhar individualmente. Para não haver outras interpretações devido ao efeito do tempo, havia um grupo de controlo que trabalhava nas mesmas condições do grupo experimental, com a excepção de não participar na 2ª etapa.

Uma das hipóteses era que a concertação atenuaria as divergências. Foram comparadas as diferenças entre as divergências nos dois grupos, na 3ª etapa e na 1ª etapa. Os resultados foram claros, a concordância entre as notas e entre as posições relativas dos alunos foi mais elevada no grupo que tinha participado numa concertação do que no grupo controlo.

---

<sup>17</sup> A validade ecológica é aqui entendida como o grau em que a pesquisa emula o mundo real (Thomas et al., 2007).

Tal como na experiência de Bonniol, Caverni e Noizet, o nosso estudo também decorreu em três etapas.

Quadro 3. Plano de pesquisa.

	1ª Etapa	2ª Etapa	3ª Etapa
Grupo de Controlo	Tarefa individual:  Classificar os jogadores do conjunto de jogos A (jogos 1, 2 e 3)	Tratamento: -«Placebo»	Tarefa individual:  Classificar os jogadores do conjunto de jogos B (jogos 4, 5 e 6)
Grupo Experimental 1		Tratamento: -Instrução para avaliar utilizando os níveis do programa (I, E e A)	
Grupo Experimental 2		Tratamento: -Instrução para avaliar utilizando os níveis do programa (I, E e A) -Instrução para avaliar identificando o jogo antes de classificar os jogadores	

Na 1ª etapa todos os observadores/avaliadores<sup>18</sup> classificaram, cada um deles isoladamente, o mesmos alunos/jogadores, num conjunto de três jogos. Na 2ª etapa os observadores/avaliadores dos grupos experimentais tiveram instrução sobre como avaliar tomando como referência os níveis de especificação do programa de andebol, nomeadamente, o nível introdução, o nível elementar e o nível avançado. Num dos grupos experimentais recomendou-se, ainda, a utilização de uma técnica de observação/avaliação que parte do jogo para o jogador. Na 3ª etapa, os mesmos observadores/avaliadores voltaram a classificar os mesmos alunos/jogadores num novo conjunto de três jogos. Quer na 1ª etapa, quer na 3ª etapa, os alunos/jogadores eram sempre diferentes de jogo para jogo e de um conjunto de jogos, o conjunto A (jogos 1, 2 e 3) para o novo conjunto de jogos, o conjunto B (jogos 4, 5 e 6).

Como as avaliações individuais incidiam sobre os desempenhos dos mesmos jogadores, era possível estudar pelos resultados da 1ª etapa, a importância inicial dos erros de classificação<sup>19</sup> e das divergências entre os avaliadores e pelos resultados da 3ª a

<sup>18</sup> Conteúdo da tarefa do avaliador: observando alunos em situação de jogo, o sujeito escolhia o nível de jogo (I, E, A, CA) que representava o desempenho de cada um dos jogadores em presença, registando-o numa ficha de observação.

<sup>19</sup> Erro de classificação: falha ao discriminar e codificar a representação do valor de um desempenho relativamente ao padrão ou critério.

importância final dos erros de classificação e das divergências e ver assim, se a 2ª tinha contribuído para atenuar esses erros e divergências. Importava ver se o facto de terem aprendido a avaliar tomando como referência os níveis do programa, usando critérios de avaliação e indicadores de observação, e, ainda, se a utilização da técnica de avaliação/observação, preconizada num dos grupos, tinha transformado suficientemente os observadores/avaliadores para que os seus erros e as suas divergências fossem menos importantes quando de novo fossem trabalhar individualmente. Também aqui, para não haver outras interpretações, havia um grupo de controlo que trabalhava nas mesmas condições dos grupos experimentais, com excepção de não participar na 2ª etapa.

Procurou-se que a única explicação para a mudança no número de erros, a variável dependente, estivesse no modo como os participantes foram tratados no episódio de instrução para avaliar, variável independente.

Sendo necessário verificar se há equivalência nos resultados da pré-testagem, pelo menos na dimensão que é medida (Sousa, 2005), procurou-se comparar os grupos intactos, que são uma dificuldade potencial no controlo da distorção da selecção e da mortalidade experimental, tomando como base os resultados da 1ª etapa e os resultados de quaisquer variáveis de controlo potencialmente relevantes (Tuckman, 2000), tais como, o estatuto profissional, anos de serviço, idade, género, ou especialidade.

Para detectar as diferenças entre os grupos utilizou-se um desenho de pesquisa *quasi* experimental, designada como de grupo de controlo não equivalente, cujos procedimentos são os mesmos do design experimental do grupo de controlo com pré-teste e pós-teste, excepto o processo de selecção aleatória dos sujeitos (Tuckman, 2000; Thomas et al., 2007).

Neste tipo de desenho de pesquisa procura-se uma maior correspondência com o ambiente do mundo real, em que os grupos não podem ser formados aleatoriamente, ao mesmo tempo que se controla o maior número possível de ameaças à validade interna (Thomas et al., 2007).

Para melhor descrever o desenho de pesquisa de grupo de controlo não-equivalente, utiliza-se a simbologia de Campbell e Stanley que foi adoptada pela generalidade dos investigadores em educação (Sousa, 2005):

O1	X1	O2
-----		
O3	X2	O4
-----		
O5		O6

A expectativa era que os grupos não apresentassem diferenças no pré-teste, o conjunto de jogos A (jogos 1, 2 e 3), mas já apresentassem mudanças do pré-teste para o

pós-teste, formado pelo conjunto de jogos B (jogos 4, 5 e 6) e que essa mudança fosse diferente para os sujeitos dos grupos experimentais e de controlo.

### **2.3.1. O estudo-piloto**

Antes da recolha definitiva de dados realizou-se um estudo-piloto com o propósito de verificar se a técnica de recolha de dados e todos os procedimentos funcionavam de acordo com o especificado para o tipo de participante previsto.

Tal como na experiência definitiva, o estudo-piloto também decorreu em três etapas. No entanto, é de referir que no estudo-piloto houve apenas um grupo experimental e no decurso da 2ª etapa o grupo de controlo não teve qualquer tipo de formação, avançando, imediatamente, para o visionamento do conjunto de jogos B (jogos 4, 5 e 6).

O estudo-piloto foi realizado no decorrer duma acção de formação sobre avaliação nos JDC realizada na Universidade Lusófona de Humanidades e Tecnologias, fazendo a experiência parte integrante do programa de formação. Utilizou-se, portanto, uma amostra propositada formada pelo grupo já constituído para participar na acção de formação.

O grupo de participantes foi constituído por professores de EF de várias escolas da região de Lisboa e estagiários de EF da Universidade Lusófona de Humanidades e Tecnologias, totalizando 129 adultos. Constituíram-se dois grupos que, em função do tipo de tratamento que tiveram, foram designados de grupo de controlo e grupo experimental. O grupo de controlo com 73 adultos e o grupo experimental com 56 adultos.

Neste estudo, também, utilizámos uma técnica de recolha de dados baseada num processo de observação sistemática em vídeo, em que os avaliadores categorizam, segundo critérios, o desempenho dos alunos/jogadores nos jogos observados. Essa avaliação foi anotada numa ficha de registo da observação (*cf.* Apêndice 1),

A principal diferença em relação ao estudo definitivo é que, em primeiro lugar, os avaliadores/observadores registavam o nível do jogo em presença. Esse registo era feito de acordo com as seguintes notações<sup>20</sup>:

I significa “Introdução”: jogo que não cumpre o nível introdução, por isso, os alunos devem aprender este nível do programa

E significa “Elementar”: jogo que cumpre o nível introdução, por isso, os alunos devem aprender o programa de nível elementar.

A significa “Avançado”: jogo que cumpre o nível elementar, por isso, os alunos devem aprender o programa de nível avançado.

CA significa “Cumpe Avançado”: jogo que cumpre o nível avançado.

---

<sup>20</sup> Notação: registo em código num sistema de classificação que representa características relevantes do objecto em apreciação.



Em segundo lugar, os avaliadores/observadores registavam os alunos que, por serem de outro nível, se destacavam do nível de jogo observado.

Os procedimentos e critérios adoptados na gravação dos jogos em vídeo e as sessões em que os dados foram recolhidos foram, em tudo, muito semelhantes às do estudo definitivo. As principais diferenças deveram-se ao facto da selecção de jogos ter sido feita a partir de um conjunto de jogos pertencentes a uma videoteca pessoal (*cf.* Anexo em CD), do questionário inicial (*cf.* Apêndice 1) não conter especificação para identificar professores orientadores de estágio, e, de no decurso da 2ª etapa o grupo de controlo não ter qualquer tipo de formação tendo avançado, imediatamente, para o visionamento do conjunto de jogos B (jogos 4, 5 e 6).

Conclusões principais:

1 - A recolha de dados não funcionou de acordo com o pretendido sendo impossível tratar os dados com o rigor necessário.

Em muitos casos, quando os observadores destacavam jogadores, esses destaques não eram coerentes com a classificação do nível de jogo.

Por vezes os jogadores destacados eram dados como exemplo do nível de jogo, não se percebendo se os outros jogadores, os não destacados, pertenciam ou não a esse nível. Outras vezes, por omissão, não era claro qual o nível de jogo dos jogadores destacados porque os observadores não acompanhavam esse destaque com a anotação do nível de jogo.

Ficou por perceber-se se os participantes entenderam que os destaques não eram para ser feitos por comparação. Por exemplo, um jogador ser melhor que os outros não é suficiente para ser destacado, a não ser que esse desempenho corresponda a outro nível de jogo.

Também não ficou claro que os participantes tenham percebido que as notações correspondiam a programa a aprender e não a programa cumprido. Ou seja, a anotação «Elementar» não significa que o jogador cumpre o nível elementar, mas sim que está apto a aprender esse nível.

Ora um dos motivos mais fortes para tal ter acontecido é que o protocolo de observação e registo pode não ter sido completamente compreendido pelos observadores sendo, por isso, necessário explicá-lo melhor. Foi decidido também utilizar-se uma ficha de registo diferente, em que tudo o que fosse anotado não estivesse sujeito a qualquer interpretação.

2 – A ficha de registo induzia, logo, uma técnica de avaliação que se traduzia em classificar em primeiro lugar o jogo e só depois os jogadores. Como isso não faz parte nem das práticas avaliativas dos professores nem da sua formação para avaliar, tornava-se

pertinente tentar perceber se essa diferença na forma de avaliar poderia produzir, ou não, classificações mais válidas. Assim, decidiu-se retirar a classificação do nível de jogo da ficha de registo e constituir mais um grupo experimental que para além da instrução sobre os níveis do programa, também, tivesse instrução sobre como classificar partindo do jogo para o jogador.

3 – O tratamento dado ao grupo experimental poderia ser mais forte se utilizasse como exemplo os jogos e os jogadores participantes no conjunto de jogos A (jogos 1, 2 e 3). Evidenciar as suas características poderia garantir uma melhor compreensão dos critérios de avaliação e dos indicadores de observação.

Tratamentos mais fortes reduzem a dispersão dos resultados de cada sujeito, tornando a variância menor, e fazem aumentar a diferença entre a média dos grupos de controlo e dos grupos experimentais, permitindo aumentar o poder dos testes (Thomas et. al., 2007).

4 – Encontrar um grupo de professores que em conjunto com o investigador pudessem constituir o chamado «padrão de ouro» (o que foi relatado já no ponto 2.2.1).

O facto dos professores do agrupamento de educação física da ESDLG reunirem condições para se constituírem como «padrão de ouro» apresentava, ainda, a vantagem de poderem participar na selecção dos alunos e dos jogos a gravar.

5 – No jogo 6 do estudo-piloto, quatro jogadores poderiam ser considerados «jogadores de fronteira», ou seja, as características evidenciadas durante o período de jogo tornavam muito difícil determinar qual o seu nível de jogo. Decidiu-se não utilizar alunos/jogadores com estas características aquando da gravação dos jogos que iriam fazer parte do estudo definitivo.

Outras conclusões:

6 - Não é necessário novo visionamento de 2 minutos antes dos jogos do conjunto B. É suficiente a adaptação às condições de filmagem (luminosidade, ângulo de filmagem, equipamentos e outros), feita antes do início do visionamento dos jogos do conjunto A.

7 – O tempo de jogo nos dois conjuntos de jogos deve estar uniformizado, o que não acontecia nos jogos do estudo-piloto em que os jogos do conjunto B tinham em média menos dois minutos que os jogos do conjunto A.

Tomando como referência as conclusões do estudo-piloto, passamos, de seguida, a descrever a forma como foi realizado o estudo.

## 2.4. Participantes

Neste estudo participaram professores de educação física de várias escolas da região de Lisboa e da região do Algarve e estagiários de educação física da Universidade Lusófona de Humanidades e Tecnologias.

Na perspectiva de procurar uma correspondência com o ambiente do mundo real da formação de professores e, também, com a prática de avaliação de produtos escolares, a selecção dos participantes assentou em quatro pressupostos:

1 - Se “a avaliação dos produtos escolares é um comportamento exercido no quadro de uma profissão” (Noizet & Caverni, 1985, p. 77), então, teriam de ser professores ou futuros professores da disciplina a participar na investigação.

2 – Se os tratamentos têm de ser capazes de produzir as alterações esperadas, então, os participantes teriam de estar no nível «certo» (Thomas et al., 2007). Considerámos no nível «certo» os professores de EF que por inerência dessa função já tinham experiência no exercício da actividade de avaliação e professores estagiários de EF que de acordo com a formação académica tida até ao momento, estavam em condições de iniciar o exercício da actividade de avaliação.

3 – Se o comportamento de avaliação é susceptível de aprendizagem, então, a formação dos professores deve incluir uma formação para a avaliação (Noizet & Caverni, 1985). Considerámos que os participantes teriam de estar envolvidos em processos de formação organizados em torno de temas relacionados com a avaliação em EF, em geral, ou com a avaliação dos JDC, em particular. Esses processos poderiam estar organizados sobre a forma de colóquios, seminários ou acções de formação, fazendo a experiência parte integrante do programa de formação.

4 – Se era necessário efectuar uma análise quantitativa dos dados recolhidos, então, o tamanho da amostra teria de ter 30 sujeitos como número mínimo para se efectuar essa análise (Sousa, 2005). Considerámos necessário garantir que cada grupo experimental ou o grupo de controlo teriam 30 sujeitos no mínimo.

Considerados estes pressupostos, tornava-se impossível atribuir as pessoas a grupos de modo aleatório. De acordo com Thomas et al. (2007), como os participantes se inscrevem de acordo com critérios pessoais, estaríamos a reduzir o seu número se lhes pedíssemos que mudassem para outro horário ou para outro lugar. Isso enfraqueceria a participação, e por isso, a mortalidade experimental.

Como os grupos seriam formados a partir de outra base que não a da formação aleatória, tratava-se, então, de constituir uma amostra «boa o suficiente» (Thomas et al., 2007) para que os quatro pressupostos enunciados anteriormente, e que presidiam à selecção dos participantes, fossem considerados satisfeitos. Assim, constitui-se, uma

amostra propositada (Sousa, 2005) com sujeitos a reunirem as características indispensáveis para se poder realizar o estudo.

#### *Caracterização da amostra*

Todos os professores seleccionados participaram em sessões que estavam integradas em acções de formação, simpósios ou seminários sobre avaliação em EF ou, mais especificamente, sobre avaliação e ensino nos JDC. Três sessões realizaram-se em Lisboa, duas na Universidade Lusófona de Humanidades e Tecnologias, uma na Escola Preparatória Marquês de Alorna, e uma quarta sessão realizou-se em Portimão, no Instituto Superior Manuel Teixeira Gomes.

Em Lisboa, na Universidade Lusófona de Humanidades e Tecnologias, num seminário de preparação do ano de estágio, sobre avaliação em Educação Física, participaram 89 professores estagiários.

Em Lisboa, na Universidade Lusófona de Humanidades e Tecnologias, numa acção de formação sobre ensino e avaliação nos JDC, participaram 32 professores.

Em Lisboa, na Escola Preparatória Marquês de Alorna, numa acção de formação sobre avaliação em EF, participaram 22 professores.

Em Portimão, no Instituto Superior Manuel Teixeira Gomes, num simpósio sobre avaliação em EF, participaram 23 professores.

Depois de explicado como os direitos e privacidade seriam protegidos, todos os professores envolvidos nestas acções acederam, voluntariamente, a participar neste estudo.

A amostra de observadores foi seleccionada de entre o conjunto dos 166 professores envolvidos nestas acções.

151 Professores, cujos dados recolhidos estavam completos, puderam fazer parte da amostra final.

Por não terem completado as tarefas de avaliação, foram retirados da amostra os professores que não tenham observado qualquer um dos seis jogos que faziam parte do estudo. Estiveram nesta circunstância aqueles que por vários motivos se ausentaram mais cedo das sessões ou estiveram momentaneamente ausentes durante o visionamento de algum jogo.

Identificámos algumas variáveis que pudessem ser responsáveis por influenciar resultados e factores de distorção na selecção. Como da composição da amostra, faziam parte alguns participantes envolvidos no processo do estágio da formação inicial de professores, foram considerados, quanto ao estatuto profissional, três tipos diferentes de professor: o professor estagiário, o professor orientador de estágio e o professor não directamente envolvido no processo de estágio.

Na mesma perspectiva, também fizemos a distinção entre especialistas e não especialistas em JDC. Foram considerados especialistas em JDC todos os participantes que no inquérito inicial revelaram ter, cumulativamente, curso de treinador de nível 1, 2 ou 3 da respectiva federação desportiva, nas modalidades de Andebol, Basquetebol, Futebol, Voleibol, Rugby, Corfebol ou Hóquei em Patins e, anos de prática como treinador.

Na tentativa de controlar ameaças à validade interna procurou-se que os participantes dos grupos de controlo e experimentais fossem o mais semelhante possível, já que os efeitos das variáveis de controlo são, também, neutralizados quando se uniformizam os diferentes grupos (Tuckman, 2000, p.130). Assim, nas sessões em que as condições o permitiram, combinámos conjuntos de participantes iguais quanto a alguma das características anteriormente referidas, e depois, atribuímos cada um deles a um grupo diferente de modo aleatório.

Foi o caso das sessões que decorreram na Universidade Lusófona de Humanidades e Tecnologias. Nessas sessões, foi possível a constituição de diferentes grupos (controlo, experimental 1 e experimental 2) porque foram reunidas condições relacionadas com a duração da formação, o número de formadores e, especialmente, um número suficiente de observadores/avaliadores. Depois de uma rápida triagem aos inquéritos previamente preenchidos, organizou-se a composição dos três grupos de forma equilibrada, quanto ao género, anos de serviço, ser ou não especialista em JDC, ou estatuto profissional.

Nas sessões que decorreram na Escola Preparatória Marquesa de Alorna e no Instituto Superior Manuel Teixeira Gomes, os participantes fizeram todos parte de grupos experimentais.

Para o presente estudo, constituíram-se três grupos que, em função do tipo de tratamento que tiveram, foram designados de grupo de controlo, grupo experimental 1 e grupo experimental 2.

A composição e caracterização da amostra final são aquelas que seguidamente apresentamos:

Participaram no estudo 151 adultos com idades entre os 21 e os 57 anos ( $M=32.81$ ,  $DP=10.19$ ) divididos em 3 grupos (ver Tabela 1):

Grupo de controlo constituído por 39 adultos (25 homens e 14 mulheres), com idades entre os 21 e os 51 anos ( $M=29.67$ ,  $DP=9.01$ ), e anos de serviço entre os 0 e os 27 anos ( $M=4.24$ ,  $DP=8.15$ ), sendo 6 (15.4%) professores, 28 (71.8%) estagiários e 5 (12.8%) professores orientadores de estágio, de entre os 39 sujeitos, 9 (23.1%) são especialistas em JDC e 30 (76.9%) não especialistas em JDC.

Grupo experimental 1 constituído por 58 adultos (34 homens e 24 mulheres), com idades entre os 21 e os 56 anos ( $M=34.74$ ,  $DP=9.88$ ), e anos de serviço entre os 0 e os 34 anos ( $M=9.05$ ,  $DP=11.46$ ), sendo 26 (44.8%) outros professores, 28 (48.3%) estagiários e 4 (6.9%) professores orientadores de estágio, de entre os 58 sujeitos, 17 (29.3%) são especialistas em JDC e 41 (70.7%) não especialistas em JDC.

Grupo experimental 2 constituído por 54 adultos (29 homens e 25 mulheres), com idades entre os 21 e os 57 anos ( $M=33.00$ ,  $DP=10.93$ ), e anos de serviço entre os 0 e os 33 anos ( $M=8.65$ ,  $DP=10.25$ ), sendo 22 (40.7%) outros professores, 27 (50.0%) estagiários e 5 (9.3%) professores orientadores de estágio, de entre os 54 sujeitos, 13 (24.1%) são especialistas em JDC e 41 (75.9%) não especialistas em JDC.

Tabela 1. Estatística descritiva da idade e anos de serviço e análise de frequência do género, estatuto profissional e especialistas em JDC dos grupos de controlo, experimental 1 e experimental 2.

Grupo	N	Idades			Anos de serviço				Tipo de professor			Especialista em JDC		
		(Género)	Min	Max	M	DP	Min	Max	M	DP	Professor	Orientador	Estagiário	Sim
Controlo	39 (25♂;14♀)	21	51	29.67	9.01	0	27	4.24	8.15	6 (15.4%)	5 (12.8%)	28 (71.8%)	9 (23.1%)	30 (76.9%)
Experimental 1	58 (34♂;24♀)	21	56	34.74	9.88	0	34	9.05	11.46	26 (44.8%)	4 (6.9%)	28 (48.3%)	17 (29.3%)	41 (70.7%)
Experimental 2	54 (29♂;25♀)	21	57	33.00	10.93	0	33	8.65	10.25	22 (40.7%)	5 (9.3%)	27 (50.0%)	13 (24.1%)	41 (75.9%)

## **2. 5. Técnica de recolha de dados**

No sentido das conclusões do estudo serem consideradas válidas, organizámos a experiência de forma a representar as condições em que os professores habitualmente avaliam os alunos. Actualmente, é muito utilizada a tecnologia de vídeo no ensino e na avaliação em EF. Gréhaigne et al. (2010) consideram que o uso do vídeo pode permitir recordar e reconhecer o que se adquiriu com a prática, tornando mais operacional a sua reutilização nas situações de jogo. Darden (1999) também analisou o uso do vídeo como ferramenta de apoio ao ensino, principalmente em performances no teatro, instrução de sala de aula e desenvolvimento da habilidade motora. Este autor considera que a importância da integração da tecnologia de vídeo em EF está em duas áreas, no feedback de instrução e na avaliação.

Neste estudo, utilizámos uma técnica de recolha de dados baseada num processo de observação sistemática em vídeo, em que nos jogos observados os avaliadores categorizam, segundo critérios, o desempenho dos alunos/jogadores. Essa avaliação, foi registada numa ficha de registo da observação (*cf.* Apendíce 3), segundo uma escala construída de modo a representar a categoria comportamental do aluno/jogador, ou seja, o seu grau de competência prática de interpretação dos conteúdos de cada nível do programa.

A escala classificatória é frequentemente utilizada na avaliação de desempenhos que implicam uma avaliação subjectiva baseada em critérios e que não se prestam a testes de habilidade objectivos (Thomas et. al., 2007). Quando organizados em grupos numéricos, não só facilita o registo dos desempenhos, como permite a sua análise estatística, (Sousa, 2005).

Assim, os avaliadores/observadores registavam a classificação de cada aluno, de acordo com as seguintes notações, equivalentes às quatro categorias de desempenho possíveis, a saber:

I significa “Introdução”: aluno que deve aprender o programa de nível introdução.

E significa “Elementar”: aluno apto a aprender o programa de nível elementar porque já cumpre o de nível introdução.

A significa “Avançado”: aluno apto a aprender o programa de nível avançado porque já cumpre o de nível elementar.

CA significa “Cumpe Avançado”: aluno que já cumpre o programa de nível avançado.

Para que as avaliações feitas pelos observadores/avaliadores pudessem ser comparadas, o investigador e o GEF da ESDLG foi considerado o «padrão de ouro». De



acordo com Kupermintz (2004), “expert judges may serve as a gold standard against which more typical judges performance can be compared and evaluated” (p.194). Nesta linha, considerámos aceitável e correcto que o acordo objectivo que verificámos no júri de especialistas constituísse o «padrão de ouro», de modo que, para efeitos deste estudo, todas as avaliações não concordantes com esse padrão foram consideradas «erros do observador».

Como medidas da variável dependente, foram contabilizadas, para cada observador, a soma de erros de avaliação nos jogos do conjunto A (jogos 1, 2 e 3), correspondentes ao pré-teste (1ª etapa), e a soma de erros de avaliação nos jogos do conjunto B (jogos 4, 5 e 6), correspondentes ao pós-teste (3ª etapa).

## **2. 6. Procedimentos**

### *A) As sessões:*

*O questionário inicial e a divisão dos formandos pelos grupos experimental 1, experimental 2 e de controlo*

No início de cada sessão, os participantes preencheram um questionário (cf. Apêndice 2) com perguntas sobre a idade, o género, a data de final de curso, o tipo de habilitação académica e profissional, o estatuto profissional (estagiário, orientador de estágio ou outro professor), o tipo de formação ou curso na área do treino desportivo em JDC, e um pequeno sumário de currículo nos JDC (curso de treinadores, anos de exercício como treinador, anos de prática como jogador, árbitro, equipas do desporto escolar, etc.).

Este questionário, permitia recolher dados sobre variáveis de interesse para o estudo, particularmente, variáveis que poderiam influenciar os erros de classificação produzidos pelos observadores.

Nas sessões que decorreram na Universidade Lusófona de Humanidades e Tecnologias, baseámo-nos numa rápida análise de parte da informação contida no questionário, para combinar conjuntos de participantes iguais em alguma característica que foram atribuídos a grupos diferentes de modo aleatório, constituindo-se, assim, os grupos experimentais e o grupo de controlo.

### *O programa da sessão e o protocolo de observação e registo*

Depois de divididos pelos três grupos do estudo (controlo, experimental 1 e experimental 2) os formandos eram informados e esclarecidos sobre a forma como iria decorrer a sessão e sobre questões relacionadas com o protocolo de observação, a saber:

1º Que iriam visionar jogos de Andebol, em vídeo, e classificar os jogadores participantes.

2º Que todos os jogos se realizavam no mesmo espaço (campo de Andebol da ESDLG) e que todos os jogadores (alunos) sendo diferentes em cada jogo, estavam identificados com um número e integravam a equipa verde ou a equipa laranja.

3º Que, em primeiro lugar, iriam observar em vídeo um jogo de Andebol de cerca de 2 minutos, para se ambientarem ao ângulo de filmagem, luminosidade e equipamentos (coletes verdes e laranjas, números, etc.).

4º Que, em segundo lugar, iriam observar em vídeo um conjunto de três jogos de Andebol (jogo 1, 2, 3), todos diferentes, de cerca de 8 minutos cada um. Numa ficha de registo (*cf.* Apêndice 3), teriam de registar, à frente do número de cada jogador, o nível de jogo (Introdução, Elementar, Avançado, Cumpre Avançado) de cada um dos jogadores de campo.

5º Que os guarda-redes não eram objecto de classificação.

6º Que esse registo poderia ser feito durante o visionamento ou nos dois minutos seguintes ao final de cada jogo. Como ocorriam entre a observação de dois jogos, esses dois minutos serviam para corrigir alguma notação e para descansar do período de visionamento anterior.

7º Que durante o período de visionamento e de registo não era permitido consultar qualquer documento ou haver comunicação entre observadores.

8º Que após o visionamento dos jogos 1, 2 e 3, haveria uma formação sobre avaliação nos JDC (30 minutos).

9º Que, finalmente, observariam em vídeo outro conjunto de três jogos de Andebol (jogos 4, 5, 6), todos diferentes, de cerca de 8 minutos cada um, tendo que utilizar todos os procedimentos adoptados durante o visionamento dos três primeiros jogos.

Antes de começar foi claramente definido o significado das notações:

Introdução – aluno que deve aprender o programa de nível introdução.

Elementar – aluno apto a aprender o programa de nível elementar porque já cumpre o de nível introdução.

Avançado – aluno apto a aprender o programa de nível avançado porque já cumpre o de nível elementar.

Cumpre Avançado – aluno que já cumpre o programa de nível avançado.

As sessões decorreram de acordo com as informações e esclarecimentos que foram prestados aos formandos, quer sobre o programa da sessão, quer sobre o protocolo de avaliação e registo.

*B) O tratamento dado a cada um dos grupos durante o episódio de instrução:*

Como recomendam Noizet e Caverni (1985), todos os professores que participaram ignoravam o motivo principal do estudo. Pretendia-se não correr o risco de deformar os resultados.

Também, todos os participantes ignoravam que tipo de tratamento estavam a receber, se o tratamento do grupo de controlo ou o tratamento dos grupos experimentais, aquilo que se designa de arranjo cego (Thomas et. al., 2007). Com este procedimento, procurava-se evitar uma maior atenção e empenho, motivados pelo facto de saberem estar a ser submetidos a uma situação experimental, o chamado efeito Hawthorne (Sousa, 2005).

*Grupo de Controlo*

De acordo com Tuckman (2000), para se poder controlar factores como o tempo, o envolvimento na experiência e a exposição aos materiais, deve-se usar o método da constância, que consiste em submeter o grupo de controlo a uma experiência paralela ao tratamento, que proporcione a mesma quantidade de exposição e o mesmo grau de envolvimento.

Assim, este grupo teve uma instrução considerada irrelevante para produzir menos erros de classificação no conjunto de jogos B (jogos 4, 5 e 6), do que aqueles que produziu no conjunto de jogos A (jogos 1, 2 e 3). Embora merecesse a mesma atenção da parte dos formadores, na verdade, recebeu um tratamento falso que não teve relação nenhuma com o desempenho na variável dependente, ou seja, um tratamento designado de placebo (Thomas et. al., 2007).

Foram abordadas questões pertinentes (*cf.* Anexo em CD), mas apenas relacionadas com o ensino e não com a avaliação nos JDC, nomeadamente:

1- A forma como muitos professores planeiam as aulas e o processo de instrução, provoca uma espécie de curto-circuito na actividade dos alunos, que impede progressos na aprendizagem.

2- Os alunos passam, sistematicamente, do treino das habilidades técnicas, em situação de exercício analítico, para o jogo formal, por vezes demasiado adulto, em espaços e com um número de jogadores que torna o contexto demasiado complexo.

3- O treino das habilidades técnicas descontextualizadas torna-se muito aborrecido e desprovido de significado para os alunos.

4- No jogo formal, por um lado, o professor comporta-se, muitas vezes, como uma espécie de polícia sinaleiro, arbitrando e dando indicações sobre o que os jogadores devem fazer, e, por outro lado, o comportamento dos alunos revela uma prática muito anárquica e demonstrativa de pouca compreensão do jogo, mesmo nos alunos mais habilidosos.

5- Aconselhou-se a que os professores escolham situações de aprendizagem baseadas em situações de jogo modificadas, para pôr em evidência problemas detectados na situação de jogo que é referência de um determinado nível.

6- Aconselhou-se que as situações de exercício analítico só devam ser usadas para resolver problemas de execução que não conseguem ser resolvidos em situação de jogo.

#### *Grupo Experimental 1*

A formação deste grupo foi orientada pelo investigador, dela constando os seguintes pontos (*cf.* Anexo em CD):

1- Depois de distribuído um exemplar do programa de Andebol, os participantes leram-no, individualmente.

2- O investigador explicou as características e especificidades de cada nível do programa.

3- Definiram-se critérios de avaliação para distinguir os diferentes níveis de jogo.

4- Definiram-se indicadores de observação que permitissem revelar um determinado nível de jogo.

5- Compararam-se os jogos (1, 2 e 3) e os jogadores visionados, anteriormente, com os níveis do programa, particularmente, os que apresentavam características que possibilitassem um melhor esclarecimento de dúvidas. Desta comparação não constou qualquer novo visionamento de imagens.

#### *Grupo Experimental 2*

A formação deste grupo foi, também, orientada pelo investigador, dela constando os mesmos cinco pontos do grupo experimental 1 (*cf.* Anexo em CD):

1- Depois de distribuído um exemplar do programa de Andebol, os participantes leram-no, individualmente.

2- O investigador explicou as características e especificidades de cada nível do programa.

3- Definiram-se critérios de avaliação para distinguir os diferentes níveis de jogo.

4- Definiram-se indicadores de observação que permitissem revelar um determinado nível de jogo.

5- Compararam-se os jogos (1, 2 e 3) e os jogadores visionados, anteriormente, com os níveis do programa, particularmente, os que apresentavam características que possibilitassem um melhor esclarecimento de dúvidas. Desta comparação não constou qualquer novo visionamento de imagens.

E ainda:

6- Recomendou-se a utilização de uma técnica de observação/avaliação e registo, que parte do jogo para o jogador. Em primeiro lugar, os observadores avaliam o jogo, situando-o num nível, e, só depois, situam cada um dos jogadores.

Como os jogadores são responsáveis pela produção colectiva que se observa, a maioria deles, ou pelo menos metade, deve ser situada no mesmo nível atribuído ao jogo. De seguida, destacam-se os jogadores que pertencem a outro nível de jogo.

*C) Critérios de avaliação e indicadores de observação*

*Distinguir o nível introdução do nível elementar:*

*Os critérios de avaliação*

Regras principais – início e recomeço do jogo, golos, foras, violações da área, dribles e passos, conduta para com o adversário e penalizações.

Pega e armação do braço.

Decidir e executar quando passar ou driblar.

Remate em salto, quando isolado na zona central.

Desmarcação - «Quebra do alinhamento».

Colocação, como defesa, entre a baliza e o adversário.

*Os indicadores de observação:*

Observa-se progressão no campo. A bola vai de «baliza a baliza».

Observam-se golos na sequência de remates em salto, não protagonizados.

*Distinguir o nível elementar do nível avançado:*

*Os critérios de avaliação*

Desmarcação rápida após recuperação da bola.

Colaboração na circulação de bola com compensação ofensiva.

Ultrapassar o adversário directo por exploração horizontal e vertical.

Marcação defensiva de vigilância e de controlo.

*Os indicadores de observação:*

Observa-se jogo com largura e profundidade.

Observam-se contra-ataques.

Observa-se o «trapézio ofensivo».

Observam-se muitas situações de 1x1, com fintas e remates apoiados e em suspensão.

Observa-se contacto físico.

## 2. 7. Análise dos dados

Os dados referentes aos questionários da caracterização da amostra e às fichas de registo, das avaliações feitas pelos observadores, foram passados para uma base de dados em suporte informático e submetidos aos tratamentos estatísticos.

Recorreu-se ao *Software* SPSS Statistics (v. 19.0; IBM SPSS, Chicago, IL) para executar os testes estatísticos. Consideraram-se estatisticamente significativas as diferenças entre médias cujo *p-value* do teste foi inferior ou igual 0.05. Os resultados são apresentados como média  $\pm$  desvio-padrão da média.

De acordo com os objectivos deste estudo, pretendeu-se na análise principal, comparar a diferença entre a soma de erros de classificação dos observadores dos três grupos, antes e depois da intervenção. Antes da intervenção, para verificar se os grupos eram semelhantes. Depois da intervenção, para determinar se, não só, os observadores com instrução prévia sobre os níveis de andebol dos PNEF, como também, os que para além dessa instrução identificaram o nível de jogo antes de classificar os jogadores produziam classificações mais válidas, errando menos, do que aqueles que não tinham tido esta instrução. Utilizou-se uma ANOVA one-way para avaliar a significância da diferença entre a soma de erros de avaliação dos observadores nos três grupos (controlo, experimental 1 e experimental 2), no conjunto de jogos A (jogos 1, 2 e 3) e no conjunto de jogos B (jogos 4, 5 e 6).

Por outro lado, comparou-se, também, a diferença entre a soma de erros de classificação dos observadores em variáveis como o género, o tipo de professor e o ser ou não especialista em JDC, antes e depois da intervenção. Antes da intervenção, para determinar a sua influência na soma de erros de classificação. Depois da intervenção, para determinar se, dentro de cada grupo, sujeitos a tratamentos diferentes, as variáveis apresentavam resultados distintos. Porque apresentavam um número muito reduzido de sujeitos, os orientadores de estágio foram integrados, quanto ao tipo de professor, na variável professor. O teste *t*-Student foi utilizado para avaliar a significância da diferença entre a soma de erros de classificação dos observadores nas variáveis género (masculino e feminino), tipo de professor (professores e estagiários), e especialista (especialistas e não especialistas nos JDC), no conjunto de jogos A (jogos 1, 2 e 3). O teste *t*-Student foi, também, utilizado para avaliar a significância da diferença entre a soma de erros de classificação dos observadores nas mesmas variáveis, agora, dentro de cada grupo (controlo, experimental 1 e experimental 2), no conjunto de jogos B (jogos 4, 5 e 6).

Avaliámos as normalidades das distribuições e a homogeneidade de variâncias que são os pressupostos destes métodos estatísticos, ANOVA one-way e o teste *t*-Student. As normalidades das distribuições foram avaliadas com o teste de Kolmogorov-Smirnov com

correção de Lilliefors (KS) ou com o teste de Shapiro-Wilk (SW), dependendo do número de observadores que integravam cada grupo de variáveis analisadas. A homogeneidade de variâncias foi avaliada com o teste de Levene baseado na média ou na mediana, dependendo da variável dependente apresentar ou não distribuição normal.

Apesar de a variável dependente não apresentar distribuição normal nalguns grupos, considerou-se que a ANOVA one-way é robusta à violação da normalidade quando os valores de assimetria (*sk*) e kurtose (*ku*) não são muito elevados (Maroco, 2011).

Tabela 2. Variável grupo no conjunto de jogos A (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Grupo	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Controlo</b>	KS(39) = 0.141; $p = 0.050$	$sk = 1.420$ ; $ku = 2.358$	$F(2,148) = 0.071$ ; $p = 0.932$
<b>Experimental 1</b>	KS(58) = 0.114; $p = 0.057$		
<b>Experimental 2</b>	KS(54) = 0.084; $p = 0.200$		

Tabela 3. Variável género no conjunto de jogos A (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Género	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Masculino</b>	KS(88) = 0.099; $p = 0.034$	$sk = 0.748$ ; $ku = 0.840$	$F(1,149) = 1.480$ ; $p = 0.226$
<b>Feminino</b>	KS(63) = 0.098; $p = 0.200$		

Tabela 4. Variável tipo de professor no conjunto de jogos A (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Tipo de professor	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Professor</b>	KS(68) = 0.098; $p = 0.178$		$F(1,149) = 0.707$ ; $p = 0.402$
<b>Estagiário</b>	KS(83) = 0.103; $p = 0.031$	$sk = 0.193$ ; $ku = 0.678$	

Tabela 5. Variável especialista no conjunto de jogos A (jogos 1,2 e 3). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Especialista	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Especialista JDC</b>	KS(39) = 0.102; $p = 0.200$		$F(1,149) = 3.703$ ; $p = 0.056$
<b>Não Espec. JDC</b>	KS(112) = 0.067; $p = 0.200$		

Tabela 6. Variável grupo no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Grupo	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Controlo</b>	KS(39) = 0.256; $p < 0.001$	$sk = 1.080$ ; $ku = 0.854$	$F(2,148) = 3.437$ ; $p = 0.035$
<b>Experimental 1</b>	KS(58) = 0.143; $p = 0.005$	$sk = 0.743$ ; $ku = 0.577$	
<b>Experimental 2</b>	KS(54) = 0.201; $p < 0.001$	$sk = 0.997$ ; $ku = 0.825$	

Tabela 7. Variável género no grupo de controlo no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Género	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Masculino</b>	SW(25) = 0.904; $p = 0.022$	$sk = 0.987$ ; $ku = 0.636$	$F(1,37) = 0.552$ ; $p = 0.462$
<b>Feminino</b>	SW(14) = 0.837; $p = 0.015$	$sk = 1.520$ ; $ku = 2.379$	

Tabela 8. Variável tipo de professor no grupo de controlo no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Tipo de professor	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Professor</b>	SW(11) = 0.903; $p = 0.201$	$sk = 1.373$ ; $ku = 2.222$	$F(1,37) = 1.552$ ; $p = 0.221$
<b>Estagiário</b>	SW(28) = 0.856; $p = 0.001$		

Tabela 9. Variável especialista no grupo de controlo no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Especialista	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Especialista JDC</b>	SW(9) = 0.961; $p = 0.811$	$sk = 1.280$ ; $ku = 1.281$	$F(1,37) = 0.046$ ; $p = 0.832$
<b>Não Espec. JDC</b>	SW(30) = 0.851; $p = 0.001$		

Tabela 10. Variável género no grupo experimental 1 no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

Género	Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk	Valores de assimetria e kurtose	Teste de Levene
<b>Masculino</b>	SW(34) = 0.919; $p = 0.015$	$sk = 0.931$ ; $ku = 1.383$	$F(1,56) = 1.125$ ; $p = 0.293$
<b>Feminino</b>	SW(24) = 0.970; $p = 0.660$		



Tabela 11. Variável tipo de professor no grupo experimental 1 no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

<b>Tipo de professor</b>	<b>Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk</b>	<b>Valores de assimetria e kurtose</b>	<b>Teste de Levene</b>
<b>Professor</b>	SW(30) = 0.949; $p = 0.158$	$sk = 1.244$ ; $ku = 3.227$	$F(1,56) = 3.139$ ; $p = 0.082$
<b>Estagiário</b>	SW(28) = 0.882; $p = 0.004$		

Tabela 12. Variável especialista no grupo experimental 1 no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

<b>Especialista</b>	<b>Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk</b>	<b>Valores de assimetria e kurtose</b>	<b>Teste de Levene</b>
<b>Especialista JDC</b>	SW(17) = 0.882; $p = 0.034$	$sk = 1.392$ ; $ku = 2.567$	$F(1,56) = 0.074$ ; $p = 0.787$
<b>Não Espec. JDC</b>	SW(41) = 0.952; $p = 0.083$		

Tabela 13. Variável género no grupo experimental 2 no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

<b>Género</b>	<b>Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk</b>	<b>Valores de assimetria e kurtose</b>	<b>Teste de Levene</b>
<b>Masculino</b>	SW(29) = 0.756; $p < 0.001$	$sk = 1.986$ ; $ku = 4.437$	$F(1,52) = 0.466$ ; $p = 0.498$
<b>Feminino</b>	SW(25) = 0.946; $p = 0.199$		

Tabela 14. Variável tipo de professor no grupo experimental 2 no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

<b>Tipo de professor</b>	<b>Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk</b>	<b>Valores de assimetria e kurtose</b>	<b>Teste de Levene</b>
<b>Professor</b>	SW(27) = 0.909; $p = 0.021$	$sk = 0.603$ ; $ku = -0.507$	$F(1,52) = 0.033$ ; $p = 0.857$
<b>Estagiário</b>	SW(27) = 0.856; $p = 0.002$	$sk = 1.408$ ; $ku = 2.439$	

Tabela 15. Variável especialista no grupo experimental 2 no conjunto de jogos B (jogos 4,5 e 6). Resultados dos testes para avaliar a normalidade das distribuições e a homogeneidade das variâncias e valores de assimetria e kurtose nos casos de violação da normalidade.

<b>Especialista</b>	<b>Teste de Kolmogorov-Smirnov ou Teste de Shapiro-Wilk</b>	<b>Valores de assimetria e kurtose</b>	<b>Teste de Levene</b>
<b>Especialista JDC</b>	SW(13) = 0.858; $p = 0.036$	$sk = 0.541$ ; $ku = -1.150$	$F(1,52) = 0.004$ ; $p = 0.947$
<b>Não Espec. JDC</b>	SW(41) = 0.892; $p = 0.001$	$sk = 1.105$ ; $ku = 1.193$	

Apesar de não se poder assumir a homogeneidade das variâncias dos grupos (controlo, experimental 1 e experimental 2), no conjunto de jogos B (jogos 4, 5 e 6), a ANOVA é robusta à violação deste pressuposto especialmente quando, como acontece com este estudo, a dimensão dos grupos não é heterogénea, as médias não são proporcionais às dimensões dos grupos e a maior variância não é superior quatro vezes à menor variância (Maroco, 2011, p. 189-191). De acordo com Pestana e Gageiro (2003), consideram-se

grupos de dimensão semelhantes, aqueles em que o quociente entre a maior dimensão e a menor é inferior a 1.5.

O teste Games-Howell foi utilizado como teste de acompanhamento (*post hoc*) por não se verificar igualdade de variâncias entre a soma de erros de classificação dos diferentes grupos.

Para determinar a possível influência de variáveis como a idade ou os anos de serviço na soma de erros de classificação produzidos pelos observadores foi utilizada uma correlação bivariada. Assim, foi utilizado o teste de Spearman, para verificar a correlação entre a idade do observador e a soma de erros de classificação cometidos, e entre os anos de serviço e a soma de erros de classificação cometidos. Esta análise foi feita, quer no conjunto de jogos A (jogos 1, 2 e 3), quer, dentro de cada grupo, no conjunto de jogos B (jogos 4, 5 e 6).

Finalmente, para avaliar o efeito dos tratamentos na soma de erros de classificação dos observadores em cada um dos três grupos (controlo, experimental 1 e experimental 2), recorreu-se ao teste *t*-Student para amostras emparelhadas. Esta técnica é recomendada porque analisa os dados dos mesmos indivíduos em ocasiões diferentes controlando diferenças individuais entre participantes, que é a maior fonte de variação na maioria dos estudos (Thomas et. al., 2007).

## **Capítulo 3**

### **Apresentação e Discussão dos Resultados**

### **3.1. Análise das classificações no conjunto de jogos A (jogos 1, 2 e 3)**

#### **3.1.1. Das diferenças entre as variáveis de caracterização da amostra**

Não se observou um efeito estatisticamente significativo das diferentes variáveis de caracterização da amostra sobre a soma de erros de classificação dos observadores no conjunto de jogos A (jogos 1, 2 e 3), designadamente:

- a) Variável género (masculino ou feminino) ( $t(149) = -1.716$ ;  $p = 0.088$ ).
- b) Variável tipo de professor (professor ou estagiário) ( $t(149) = 1.677$ ;  $p = 0.096$ ).
- c) Variável especialista (especialista ou não especialista em JDC) ( $t(149) = 0.618$ ;  $p = 0.537$ ).

Não se observou uma correlação entre a variável idade e a soma de erros de classificação dos observadores no conjunto de jogos A (jogos 1, 2 e 3) ( $r_s = 0.155$ ;  $p = 0.058$ ), nem entre a variável anos de serviço e a soma de erros de classificação dos observadores no conjunto de jogos A (jogos 1, 2 e 3) ( $r_s = 0.079$ ;  $p = 0.342$ ).

Surpreendentemente, os resultados mostram que os professores com mais anos de serviço e idade, ou seja, potencialmente mais experientes, com estatuto profissional como é o caso dos professores e que até integravam orientadores de estágio, ou com currículo e formação para serem considerados especialistas num qualquer JDC, não se diferenciaram dos professores mais novos e com menos anos de serviço, dos estagiários ou dos não especialistas em JDC.

Era expectável que especialistas num dado JDC, com anos de prática de ensino do jogo e com conhecimentos relativos à didáctica específica dos JDC, avaliassem melhor que os não especialistas. Supunha-se que professores com vários anos de serviço, mais experientes a ensinar e a avaliar alunos, se distinguíssem dos professores inexperientes. Também era previsível que os professores, conjunto de que fazem parte orientadores de estágio, avaliassem melhor que os estagiários, sendo até verdade que os orientam, também, nesta vertente profissional.

O que querem dizer, então, estes dados surpreendentes?

Provavelmente, estes resultados querem dizer que os procedimentos utilizados e os conhecimentos mobilizados, nesta experiência, são estranhos, quer para professores experientes ou inexperientes, quer para não especialistas ou especialistas em JDC. Isto pode levar-nos a induzir a hipótese de que os professores não sabem utilizar os níveis do programa para avaliar, nem estão habituados a fazê-lo. Eventualmente, não estabelecem critérios e procedimentos de avaliação comuns, nem participam em processos de aferição de critérios para reduzir a subjectividade e as divergências de avaliação. Ficando muito isolados nas suas próprias idiossincrasias como avaliadores, não beneficiam da formação recíproca que o trabalho colectivo entre pares proporciona.

Paradoxalmente, de acordo com Carreiro da Costa e Piéron (1997), os professores consideram os PNEF uma ferramenta fundamental para a acção profissional. Segundo os autores, estudos em Portugal parecem sugerir que a maioria dos professores concordam com os PNEF porque são uma orientação para planear e ensinar; são um guia para a reflexão crítica do trabalho efectuado e reforçam a identidade profissional.

Parece, pois, que os professores embora reconhecendo a importância dos PNEF não aproveitam a possibilidade e a vantagem que é ter níveis que podem constituir-se como referência para a avaliação e não apenas para a organização de unidades didácticas.

### 3.1.2. Das diferenças entre grupos

Não se observou um efeito estatisticamente significativo da variável grupo (controlo, experimental 1 ou experimental 2) sobre a soma de erros de classificação dos observadores no conjunto de jogos A (jogos 1, 2 e 3) ( $F(2,148) = 0.062$ ;  $p = 0.940$ ).

Tabela 16. Estatística descritiva ( $M$ ,  $DP$ ) da soma de erros por conjunto de jogos (1+2+3) para os grupos de controlo, experimental 1 e experimental 2, e valores da análise de variância univariada relativa à comparação dos erros entre os grupos.

Jogos	Grupo			ANOVA	
	Controlo Erros $M$ ( $DP$ )	Experimental 1 Erros $M$ ( $DP$ )	Experimental 2 Erros $M$ ( $DP$ )	$F(2,148)$	$p$
1+2+3	8.92 (4.78)	9.14 (4.29)	9.24 (4.04)	.062	.940

A partir destes dados, foi elaborado um gráfico que representa os três grupos (controlo, experimental 1 e experimental 2) pelas médias da soma de erros de classificação.

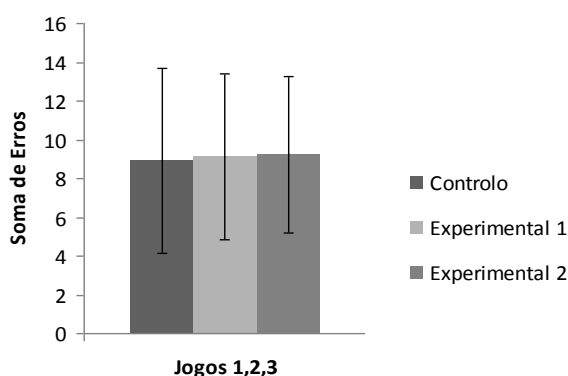


Figura 1. Média da soma de erros de classificação dos observadores dos três grupos participantes no estudo, no conjunto de jogos A (jogos 1, 2 e 3). As barras de erro representam o desvio-padrão.

Este resultado parece demonstrar que a tentativa de controlar ameaças à validade interna foi bem sucedida. Ao combinarmos em cada grupo conjuntos de participantes iguais em determinadas características como o género, os anos de serviço, o ser ou não especialista em JDC, ou o estatuto profissional e depois, atribuí-los a um grupo diferente de modo aleatório, permitiu que os participantes dos grupos de controlo e experimentais

fossem o mais semelhante possível. Essa uniformização dos diferentes grupos neutralizou os efeitos das variáveis de controlo (Tuckman, 2000, p. 130).

No momento antes da intervenção os três grupos (controlo, experimental 1 e experimental 2) parecem, pois, partir de uma base de igualdade, não se diferenciando em relação à quantidade de erros de classificação que cometem.

### **3.2. Análise das classificações no conjunto de jogos B (jogos 4, 5 e 6)**

#### **3.2.1. Das diferenças entre as variáveis de caracterização da amostra dentro de cada grupo**

##### *Grupo de controlo*

Não se observou um efeito estatisticamente significativo das diferentes variáveis de caracterização da amostra sobre a soma de erros de classificação dos observadores do grupo de controlo no conjunto de jogos B (jogos 4, 5 e 6), designadamente:

- a) Variável género (masculino ou feminino) ( $t(37) = 0.014$ ;  $p = 0.989$ ).
- b) Variável tipo de professor (professor ou estagiário) ( $t(37) = 0.030$ ;  $p = 0.976$ ).
- c) Variável especialista (especialista ou não especialista em JDC) ( $t(37) = 0.034$ ;  $p = 0.973$ ).

Não se observou uma correlação entre a variável idade e a soma de erros de classificação dos observadores do grupo de controlo no conjunto de jogos B (jogos 4, 5 e 6) ( $r_s = 0.218$ ;  $p = 0.183$ ), nem entre a variável anos de serviço e a soma de erros de classificação dos observadores do grupo de controlo no conjunto de jogos B (jogos 4, 5 e 6) ( $r_s = -0.153$ ;  $p = 0.359$ ).

##### *Grupo experimental 1*

Não se observou um efeito estatisticamente significativo das diferentes variáveis de caracterização da amostra sobre a soma de erros de classificação dos observadores do grupo experimental 1 no conjunto de jogos B (jogos 4, 5 e 6), designadamente:

- a) Variável género (masculino ou feminino) ( $t(56) = -1.725$ ;  $p = 0.090$ ).
- b) Variável tipo de professor (professor ou estagiário) ( $t(56) = 0.302$ ;  $p = 0.763$ ).
- c) Variável especialista (especialista ou não especialista em JDC) ( $t(56) = -1.831$ ;  $p = 0.072$ ).

Não se observou uma correlação entre a variável idade e a soma de erros de classificação dos observadores do grupo experimental 1 no conjunto de jogos B (jogos 4, 5 e 6) ( $r_s = 0.005$ ;  $p = 0.972$ ), nem entre a variável anos de serviço e a soma de erros de classificação dos observadores do grupo experimental 1 no conjunto de jogos B (jogos 4, 5 e 6) ( $r_s = -0.070$ ;  $p = 0.611$ ).

### *Grupo experimental 2*

Não se observou um efeito estatisticamente significativo das diferentes variáveis de caracterização da amostra sobre a soma de erros de classificação dos observadores do grupo experimental 2 no conjunto de jogos B (jogos 4, 5 e 6), designadamente:

- a) Variável género (masculino ou feminino) ( $t(52) = -1.985$ ;  $p = 0.052$ ).
- b) Variável tipo de professor (professor ou estagiário) ( $t(52) = 0.236$ ;  $p = 0.814$ ).
- c) Variável especialista (especialista ou não especialista em JDC) ( $t(52) = -0.235$ ;  $p = 0.815$ ).

Não se observou uma correlação entre a variável idade e a soma de erros de classificação dos observadores do grupo experimental 2 no conjunto de jogos B (jogos 4, 5 e 6) ( $r_s = 0.044$ ;  $p = 0.752$ ), nem entre a variável anos de serviço e a soma de erros de classificação dos observadores do grupo experimental 2 no conjunto de jogos B (jogos 4, 5 e 6) ( $r_s = 0.071$ ;  $p = 0.611$ ).

Parece verificar-se que os vários conjuntos de professores, que correspondem às variáveis analisadas (género, estatuto profissional, especialista, idade, anos de serviço), não reagiram de maneira diferente ao tratamento especificamente efectuado em cada grupo. Quer no grupo de controlo, quer nos grupos experimentais, não houve diferenças estatisticamente significativas em relação ao género (homem/mulher), estatuto profissional (professor/estagiário), ou especialista (especialista em JDC/não especialista em JDC), nem correlações envolvendo a idade ou os anos de serviço.

Tal como o verificado antes da intervenção, quando avaliaram o conjunto de jogos A (jogos 1, 2 e 3), também estas características parecem não ter feito diferença depois da intervenção, quando avaliaram o conjunto de jogos B (jogos 4, 5 e 6).

### **3.2.2. Das diferenças entre grupos**

Observou-se um efeito estatisticamente significativo da variável grupo (controlo, experimental 1 e experimental 2) sobre a soma de erros de classificação dos observadores no conjunto de jogos B (jogos 4, 5 e 6) ( $F(2,148) = 31.857$ ;  $p < 0.001$ ;  $\eta^2_p = 0.301$ ;  $\pi = 1$ ). Observou-se uma dimensão de efeito e potência de teste elevada. O grupo experimental 2 ( $M = 2.59$ ,  $DP = 2.29$ ,  $n = 54$ ) apresentou uma soma de erros de avaliação menor, seguido do grupo experimental 1 ( $M = 4.14$ ,  $DP = 2.47$ ,  $n = 58$ ) e em terceiro lugar nesta ordenação o grupo de controlo ( $M = 7.51$ ,  $DP = 4.20$ ,  $n = 39$ ). De acordo com o teste *post-hoc* Games-Howell ocorreram diferenças estatisticamente significativas entre os três grupos. Entre o grupo de controlo e o grupo experimental 1 ( $p < 0.001$ ). Entre o grupo de controlo e o grupo experimental 2 ( $p < 0.001$ ). Entre o grupo experimental 1 e o grupo experimental 2 ( $p = 0.002$ ).

Tabela 17. Estatística descritiva ( $M$ ,  $DP$ ) da soma de erros por conjunto de jogos (1+2+3 e 4+5+6) para os grupos de controlo, experimental 1 e experimental 2, e valores da análise de variância univariada relativa à comparação dos erros entre os grupos.

Jogos	Grupo			ANOVA	
	Controlo Erros $M(DP)$	Experimental 1 Erros $M(DP)$	Experimental 2 Erros $M(DP)$	$F(2,148)$	$P$
1+2+3	8.92 (4.78)	9.14 (4.29)	9.24 (4.04)	.062	.940
4+5+6	7.51 (4.20)	4.14 (2.47)	2.59 (2.29)	31.857	.000

Estes dados permitiram elaborar o gráfico da Figura 2. que representa os três grupos (controlo, experimental 1 e experimental 2) pelas médias da soma de erros de classificação.

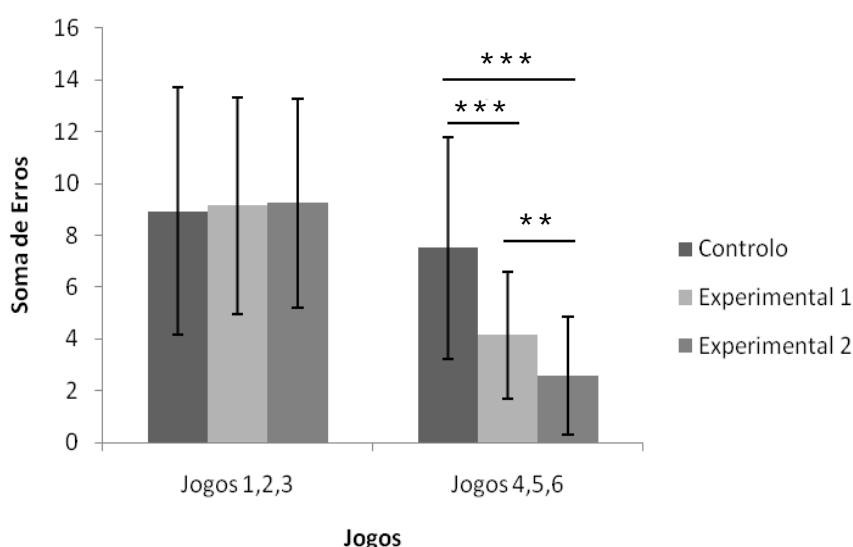


Figura 2. Média da soma de erros de avaliação dos observadores dos três grupos participantes no estudo, no conjunto de jogos A (jogos 1, 2 e 3) e no conjunto de jogos B (jogos 4, 5 e 6). As barras de erro representam o desvio-padrão. Os asteriscos representam médias significativamente diferentes de acordo com o teste Games-Howell para  $\alpha = 0.05$ .

Em relação à primeira hipótese de trabalho – “Não há diferença entre as classificações do desempenho dos jogadores em prova autêntica de andebol, produzidas, por um lado, pelos professores com instrução prévia sobre os níveis de andebol dos PNEF e, por outro lado, as classificações produzidas por professores sem essa instrução” – julgamos poder afirmar que:

Não se confirma que, professores com instrução prévia sobre os níveis de andebol dos PNEF e professores sem essa instrução, produzem classificações semelhantes.



Pelo contrário, existe uma relação entre, por um lado, o conhecimento que os professores têm dos três níveis de especificação (Introdução, Elementar e Avançado) do programa de Andebol e, por outro lado, o número de erros cometidos pelos avaliadores.

Essa relação apresentou um sentido inverso, ou seja, o conhecimento dos níveis do programa e a sua utilização na avaliação do desempenho individual em situação de jogo, fez diminuir, significativamente, o número de erros cometidos pelos avaliadores.

Nas provas abertas, como é o caso das situações de avaliação baseadas em contextos de jogo, a tendência natural para se tentar padronizar tudo o que se quer classificar é contrariada pela realidade. De facto, o jogo não é padronizável, tornando muito difícil qualquer esforço que se faça no sentido de atribuir-lhe uma pontuação. Os níveis de jogo resolvem esta dificuldade, contribuindo para que se possa avaliar provas não padronizáveis, porque permitem, precisamente, dar uma notação certa ao jogo e ao desempenho individual no jogo.

Nesta experiência, a definição dos níveis de jogo que está sistematizada nos PNEF constituiu uma referência para a avaliação, permitindo classificar os jogadores a partir do jogo. Verificou-se que os observadores que estudaram o programa passaram a conhecer melhor a especificidade da definição de cada nível. Observando os alunos em situação de jogo, foram capazes de atribuir significado aos comportamentos evidenciados, conseguindo classificar mais consistentemente o desempenho individual dos jogadores.

É de sublinhar que quanto à importância dos níveis de jogo, autores como Harvey (2006) e Tallir et al (2004) constatarem que na aplicação dos instrumentos de avaliação autêntica mais conhecidos na literatura publicada, e.g. o GPAI, faltam níveis de jogo estruturados que sirvam de referência para a avaliação. Como os autores reconhecem, esses instrumentos ficam-se por uma discriminação dos jogadores entre si, por uma comparação de quem é melhor ou pior entre os avaliados, mas não atribuem um significado, uma qualidade ao desempenho.

Ora, de acordo com Carreiro da Costa e Piéron (1997), esta definição de níveis em progressão, sistematizada nos PNEF, constitui uma excelente ajuda para os professores poderem lidar melhor com uma população escolar constituída por alunos que são muito diferentes.

Admitindo que a distinção do conteúdo dos níveis nos PNEF constitui uma sistematização válida, os níveis (I, E e A) podem também ser uma vantagem no processo de avaliação, porque como se verificou, os observadores com instrução focada nesses níveis mais facilmente situaram a qualidade do jogador no jogo, produzindo depois da instrução classificações mais consistentes do que as efectuadas anteriormente.

De acordo com Harlen (2004) a fiabilidade é maior se os professores compreenderem bem os critérios de avaliação porque ajuda-os a identificar e a apreciar as evidências que são relevantes. Também, segundo Gerard (2006), quando se utilizam instrumentos abertos de avaliação, são os indicadores que permitem operacionalizar os critérios de maneira a saber-se o que se deve exactamente observar. Foi o que realmente verificámos. Que no tratamento efectuado nos grupos experimentais, foi absolutamente fundamental, no episódio de instrução, o conhecimento e utilização de critérios de avaliação para distinguir níveis de jogo e de indicadores de observação que permitissem situar o jogo.

Por outro lado, este episódio de instrução garantiu alguma concordância entre os professores dos grupos experimentais. Um número muito apreciável de avaliadores, revelou maior consistência nas suas classificações, justamente, em provas de avaliação autênticas que não são repetíveis. Esta questão é realmente importante, já que de acordo com Gardner e Harlen (2010), a fiabilidade é baixa quando não há moderação entre os professores. Pelo contrário, quando os professores estão empenhados em tarefas comuns é possível reduzir as ameaças à fiabilidade (Harlen, 2007). Na verdade, em conjunto, os indicadores de observação e os critérios de avaliação, contribuíram para que os observadores passassem a situar os jogadores e o jogo com mais facilidade, alinhando, claramente, as suas apreciações.

Ao dominarem os níveis do programa, tomando-os como referência para a avaliação, e ao utilizarem os mesmos critérios de avaliação e indicadores de observação, os professores passaram a trabalhar numa base comum. Ficaram menos sujeitos às idiosincrasias e ao relativismo nas suas opções individuais de avaliação. De facto, como considera Wiliam (2004), o requisito principal para as avaliações em sala de aula serem válidas e fiáveis é todos os professores partilharem o mesmo constructo de qualidade, formando uma comunidade de prática.

Verificámos que, apenas com um episódio de meia hora de instrução para classificar, os resultados passaram a ser significativamente diferentes quando se compararam os professores dos grupos experimentais com os professores do grupo de controlo.

Como explicam Noizet e Caverni (1985), se dois avaliadores avaliam a mesma produção de maneira diferente é porque não recorrem ao mesmo modelo de referência. Não dispondo das mesmas informações não esperam encontrar os mesmos produtos. Para estes autores, essas flutuações não são aleatórias e o comportamento de avaliação pode ser analisado, modificado e aprendido. Sendo as capacidades técnicas dos professores, capacidades aprendidas, também avaliar exige aprendizagem,

No entanto, a aprendizagem proporcionada pelo treino de observação e avaliação e pelos processos de aferição de critérios exigem tempo disponível para se poderem realizar e uma verdadeira colaboração entre professores.

Tais práticas integram-se naquilo que Pacheco (1994), Roldão (1998, 2006) e Clímaco (2005) consideram ser necessário nas escolas, uma cultura de colaboração entre os professores. Mas, as soluções para se garantir uma boa consistência nas avaliações, como especificar as tarefas e os critérios de avaliação, moderação ou constituir uma comunidade de avaliação, implicam tempo (Harlen, 2004). As mudanças só acontecem com tempo para trabalhar em conjunto (Darling-Hammond, 2004).

Provas autênticas em matérias abertas que são mais difíceis de elaborar, aplicar e classificar (Nazario, 2005; Fernandes, 2005), ainda exigem mais tempo e mais colaboração entre pares. Como já vimos anteriormente, não só existe uma tradicional ausência de cultura de avaliação (Fernandes, 1994, 2005; Fernandes et al., 1996), como as práticas de avaliação estão marcadas pela solidão e pelo individualismo docente (Clímaco, 2005; Fernandes, 2005; Fernandes et al., 1996). É também, por isso mesmo, por não se valorizar suficientemente as práticas de trabalho colaborativo em avaliação, que não é comum ser-lhe atribuído tempo, na distribuição do serviço docente.

Então, se com meia hora de instrução conseguiram-se melhores resultados, é legítimo pensar, como expectável, que com aferição de critérios e treino de avaliação os professores possam chegar a níveis muito mais elevados de acordo e de fiabilidade das classificações. Como refere Wiggins (1990) a avaliação autêntica não evita o julgamento humano. Esse julgamento pode ser melhorado com sessões de treino de avaliação em que se utilizem exemplos de padrões de desempenho.

Em relação à segunda hipótese de trabalho – “Não há diferença entre as classificações do desempenho dos jogadores em prova autêntica de andebol, produzidas, por um lado, pelos professores com instrução prévia para identificar o nível de jogo antes de classificar os jogadores e, por outro lado, as classificações produzidas por professores sem essa instrução” – julgamos poder afirmar que:

Não se confirma que, professores com instrução prévia para identificar o nível de jogo antes de classificar os jogadores e professores sem essa instrução, produzem classificações semelhantes.

Pelo contrário, existe uma relação entre «identificar o jogo antes de classificar cada jogador» e o número de erros cometidos pelos avaliadores.

Os observadores que nos procedimentos de avaliação em primeiro lugar identificaram o jogo foram os que melhor classificaram os jogadores.

Na verdade, os resultados revelam que, quando avaliaram o conjunto B (jogos 4, 5 e 6), não só os observadores dos grupos experimentais, por comparação com os do grupo de controlo, cometeram menos erros, como também, diferenciaram-se entre si. Ou seja, os observadores do grupo experimental 2 cometeram, também, menos erros que os observadores do grupo experimental 1

Ao adoptar e utilizar este procedimento foram mais eficazes, porque em vez de estar a observar e avaliar cada aluno individualmente, avaliaram «todos ao mesmo tempo», ganhando tempo que pôde ser utilizado para esclarecer dúvidas ou verificar os destaques.

Estando, os indicadores de observação, centrados nos aspectos e nas qualidades específicas que identificam um nível de jogo, permitiram mais facilmente aos observadores perceber se os critérios de avaliação eram, ou não, cumpridos e, consequentemente, atribuir a notação certa ao jogo (I, E ou A).

Tome-se como exemplo indicadores de observação utilizados neste estudo (GEF-ESDLG, 2010). Observar um jogo de andebol sem golos quer, inevitavelmente, dizer que os alunos em presença não cumprem o nível introdução.

Como o andebol é caracterizado por ser um jogo com muitos golos, a partir do momento que se cumpre o nível introdução qualquer jogo que se observe tem sempre muitos golos, por melhor que seja a competência individual defensiva de cada jogador ou a qualidade colectiva defensiva em cada equipa. Observar-se-ão sempre muitos golos quer se trate de jogos de nível elementar ou de nível avançado, típicos de contextos escolares, quer se trate de jogos de competições federadas ou de jogos entre as melhores equipas profissionais do mundo.

Então porque é que esse indicador de observação, centrado nos aspectos globais do jogo, é útil para um observador poder dizer se o nível introdução é ou não é cumprido? Porque se os defesas, só pelo facto de existirem como tal, mesmo sem competências defensivas relevantes, forem obstáculos suficientes para os atacantes não terem sucesso, não haver golos só pode significar que os jogadores ou não progridem no campo ou não conseguem finalizar com êxito.

O que significa, então, não haver progressão no campo? Significa que, por um lado, as acções do portador da bola não são adequadas, não decidindo e executando com a qualidade necessária, quer para passar a bola a colegas com linha de passe abertas, quer para progredir em drible se não tiver oposição e, por outro lado, os colegas de equipa não se desmarcam, oferecendo linhas de passe que permitam manter a posse e a progressão da bola. E o que significa não haver finalização com êxito? Significa que os jogadores em situação isolada contra um guarda-redes, que tem um desempenho correspondente a este nível de jogo, não conseguem rematar em salto e marcar golo.

Quer dizer, olhar para um jogo sem golos significa, no mínimo, que o critério «remate em salto» não é cumprido, e caso também não haja progressão no campo, significa que critérios como «decidir e executar quando passar ou driblar» ou/e a «desmarcação», também, não são cumpridos (GEF-ESDLG, 2010).

No nível introdução o jogo pode caracterizar-se por não haver progressão no campo ou, por haver progressão no campo, indo a bola de «baliza a baliza», mas não existirem finalizações com êxito. Quer dizer, se os indicadores de observação forem bem elaborados podem não só ajudar situar o jogo em relação aos níveis do programa – golos não protagonizados significam que o jogo cumpre o nível introdução, ausência de golos significa que o jogo não cumpre o nível introdução, como, também, ajudar a determinar quais os critérios de avaliação que são ou não são cumpridos e, portanto, a identificar as causas dos problemas e das limitações.

Como já vimos, no nível introdução (Bom et al, 1989) ir a bola de «baliza a baliza» e não conseguir-se marcar golos tem um significado e não ir a bola de «baliza a baliza» já tem outro. Essa observação pode, também, ter grande relevância na avaliação formativa e permitir planear com mais coerência os passos seguintes da aprendizagem. Por exemplo, pode centrar a aprendizagem no domínio do remate em salto ou, então, nos domínios da decisão e execução de quando passar ou driblar e da desmarcação. Isso implica utilizar formas modificadas de jogo que dêem relevo ao problema identificado e que permitam muitas oportunidades de exercitação, e que, naturalmente, podem ser diferentes para a aprendizagem e aperfeiçoamento do remate em salto, daquelas que se utilizam para o treino da decisão e execução de quando passar ou driblar ou para a desmarcação.

Mas, também, na distinção entre o nível elementar e o nível avançado (Bom et al, 1989) podemos encontrar situações semelhantes. Tome-se como exemplo o indicador de observação «contacto físico» (GEF-ESDLG, 2010). Uma característica marcante do jogo de andebol é a existência de contacto físico. Essa característica é-lhe dada pelo facto dos defesas utilizarem a possibilidade concedida pela regra 8, a regra da «conduta para com o adversário», que permite que a oposição às acções do adversário possa ser feita utilizando o próprio corpo para lhe barrar o caminho. Essa possibilidade materializa-se numa técnica de marcação, designada de marcação de controlo. É o seu domínio por parte dos defensores que oferece ao jogo essa marca tão vinculada, o contacto físico permanente. Ou seja, se num jogo não observarmos contacto físico, não se pode afirmar que o jogo é de nível avançado. Quer isto dizer, do ponto de vista defensivo, que os alunos ainda não dominam competências decisivas do nível elementar.

Também, neste nível de jogo, os indicadores de observação podem revelar competências diferentes. Jogar no ataque em situações de 1X1, utilizando fintas e remates

em apoio e em suspensão para ultrapassar o adversário directo, é diferente se a oposição dominar técnicas de marcação, como a marcação de vigilância ou a marcação de aproximação ou se, para além destas técnicas, também dominar a marcação de controlo.

Repare-se, então, na estatística sumária dos seis jogos observados que corresponde a dados agrupados em itens que tomam como referência os indicadores de observação (quadros 4 a 9):

Quadro 4. Frequências de remates em salto, golos e progressões no campo no jogo 1.

JOGO 1					
Equipa verde	Remates em salto	Golos	Equipa laranja	Remates em salto	Golos
Nº 5	2	2	Nº 2	4	3
Nº 2	4	1	Nº 3	6	4
Nº 4	2	0	Nº 1	0	0
Nº 3	5	2	Nº 4	1	1
. 27 Progressões no campo (bola indo de uma baliza à outra). . 24 Acções de finalização em remate em salto, realizadas por 7 jogadores. . 13 Golos, marcados por 6 jogadores.					

Quadro 5. Frequências de remates em salto, golos e progressões no campo no jogo 2.

JOGO 2					
Equipa verde	Remates em salto	Golos	Equipa laranja	Remates em salto	Golos
Nº 2	0	0	Nº 1	0	0
Nº 3	0	0	Nº 8	8	7
Nº 5	0	0	Nº 5	0	0
Nº 8	0	0	Nº 3	0	0
. 12 Progressões no campo (bola indo de uma baliza à outra). . 9 Acções de finalização em remate em salto, realizadas por 2 jogadores. . 7 Golos, marcados por 1 jogadores.					

Quadro 6. Frequências de remates em salto, golos e progressões no campo no jogo 3.

JOGO 3					
Equipa verde	Remates em salto	Golos	Equipa laranja	Remates em salto	Golos
Nº 8	5	4	Nº 4	3	2
Nº 3	5	4	Nº 5	4	2
Nº 5	0	0	Nº 3	0	0
Nº 2	0	0	Nº 2	0	0
. 24 Progressões no campo (bola indo de uma baliza à outra). . 17 Acções de finalização em remate em salto, realizadas por 4 jogadores.					

. 12 Golos, marcados por 4 jogadores.

Quadro 7. Frequências de remates em salto, golos e progressões no campo no jogo 4.

JOGO 4					
Equipa verde	Remates em salto	Golos	Equipa laranja	Remates em salto	Golos
Nº 5	6	6	Nº 3	0	0
Nº 1	0	0	Nº 8	0	0
Nº 2	0	0	Nº 5	0	0
Nº 10	0	0	Nº 4	0	0
. 17 Progressões no campo (bola indo de uma baliza à outra). . 6 Acções de finalização em remate em salto, realizadas por 1 jogadores. . 9 Golos, marcados por 3 jogadores.					

Quadro 8. Frequências de remates em salto, golos e progressões no campo no jogo 5.

JOGO 5					
Equipa verde	Remates em salto	Golos	Equipa laranja	Remates em salto	Golos
Nº 5	1	1	Nº 4	5	4
Nº 3	3	1	Nº 5	3	3
Nº 1	1	1	Nº 2	0	0
Nº 8	3	1	Nº 7	3	3
. 24 Progressões no campo (bola indo de uma baliza à outra). . 19 Acções de finalização em remate em salto, realizadas por 7 jogadores. . 14 Golos, marcados por 7 jogadores.					

Quadro 9. Frequências de remates em salto, golos e progressões no campo no jogo 6.

JOGO 6					
Equipa verde	Remates em salto	Golos	Equipa laranja	Remates em salto	Golos
Nº 10	3	3	Nº 3	4	2
Nº 5	3	2	Nº 7	5	1
Nº 3	0	0	Nº 1	0	0
Nº 7	0	0	Nº 5	0	0
. 19 Progressões no campo (bola indo de uma baliza à outra). . 15 Acções de finalização em remate em salto, realizadas por 4 jogadores. . 9 Golos, marcados por 5 jogadores.					

Concretamente, foi isto que os elementos do grupo experimental 2 observaram nos jogos 4, 5 e 6. Sem estarem a colocar cruzes em itens de uma ficha de registo, ou a

recolherem dados para elaborar uma estatística de jogo semelhante a esta, perceberam e interpretaram o essencial do que se observava no jogo.

Utilizando os indicadores de observação conseguiram em primeiro lugar classificar o jogo e depois os jogadores sem necessitar de contar o número de vezes que um determinado comportamento era observado. Ou seja, perceberam sem ter que contar.

Perceberam que no jogo 4, um jogo de nível I, existiam poucas progressões da bola no campo (17) e, também, que as poucas acções de finalização com remate em salto (6) eram protagonizadas por um único jogador (nº5 verde), ao contrário do jogo 5, jogo de nível E, em que a bola ia frequentemente de baliza a baliza (24) e as acções de finalização estavam distribuídas pela esmagadora maioria dos jogadores, havendo apenas um jogador com fraca participação no jogo (nº2 laranja).

Perceberam no jogo 6, que pode ser considerado um «jogo de fronteira» entre o nível I e o nível E, que quatro jogadores protagonizavam todas as acções de finalização em remate em salto e que os outros quatro jogadores, mesmo quando se encontravam em situação isolada face ao guarda-redes, não rematavam em salto.

Eventualmente tiveram dificuldades em decidir qual o nível a atribuir a determinados jogadores cujos comportamentos não foram claros, ou não tiveram muitas oportunidades de os expressar, os «jogadores de fronteira». Por exemplo, os jogadores nº 1 ou nº 5 da equipa verde, que no jogo 5 só aparecem uma vez a rematar em salto.

Na verdade, o que os observadores do grupo experimental 2 fizeram foi tomar como ponto de partida os níveis de especificação da matéria e verificar quais os indicadores que revelavam um nível, sem estarem exclusivamente focados no desempenho individual. Assim, tentaram, em primeiro lugar atribuir um significado ao desempenho colectivo para, em seguida, melhor compreender cada jogador. Como refere Elias (1997), cada uma das partes que constituem um todo percebem-se e analisam-se, precisamente, a partir da compreensão do funcionamento do todo.

Por outro lado, também não estiveram a avaliar isoladamente os aspectos particulares do desempenho de cada jogador, e a somá-los para chegar a uma pontuação que reflectisse um nível de jogo.

De facto, por mais bem definidos que estejam os constructos a avaliar, como a tomada de decisão ou os que expressam a execução de técnicas necessárias ao bom desempenho no jogo, só a sua simples separação pode provocar inferências equivocadas. «Decidir bem» e «executar mal» ou «decidir mal» e «executar bem» são associações de existência improvável.

Como é que alguém pode pensar bem e tomar a decisão certa, se depois, na sequência do seu desempenho a equipa perdeu a posse de bola, porque:



- Um passe foi mal direccionado ou a sua execução não tomou em consideração as qualidades do possível receptor;
- Uma progressão em drible foi executada deficientemente, fazendo perder o controlo da bola;
- Um remate foi pouco potente e tão fácil de defender que permitiu o contra-ataque eficaz do adversário.

Ou, pelo contrário, como é que alguém pode pensar mal e tomar a decisão errada, se depois, o seu «remate impossível» deu golo ou, então o seu «passe impossível» permitiu a um colega de equipa uma recepção e finalização com sucesso.

No entanto, uma ressalva deve ser feita. Durante o processo de aprendizagem pode ser importante e necessário promover a tomada de decisões que não conduzem a um sucesso imediato. Frequentemente, pode ser o melhor para aprender arriscar fazer um remate que não dá origem a um golo ou arriscar fazer um passe ofensivo que não chega ao seu destino. Compreender que aquelas são as melhores soluções para a situação e o momento, colocará a necessidade de aperfeiçoar a execução e de treinar as capacidades motoras necessárias ao sucesso das acções em causa. É aqui que, incorporada no estudo do todo, a análise das partes ganha o seu espaço, podendo ser necessária e útil.

Durante os últimos trinta anos vários autores, preconizando diferentes perspectivas, basearam os seus modelos de ensino numa aprendizagem autêntica dos jogos (Barreto et al., 1980; Bunker & Thorpe, 1982; Siedentop, 1994; Griffin et al., 1997; Kirk & McPhail, 2002; Graça et al., 2003; Gréhaigne et al., 2005; Chow et al., 2007). Em oposição a um ensino isolado e descontextualizado das técnicas, elegeram o jogo como a referência à volta da qual se organiza todo o processo de aprendizagem. No centro das preocupações do método de ensino passou a estar a resolução dos problemas que o jogo coloca, a partir do conhecimento e domínio prático das acções tácticas e dos princípios que as orientam.

Actualmente, vários autores como Gréhaigne et al. (2005), entre outros, reconhecem a importância duma avaliação autêntica alinhada com o ensino. Harvey (2006), reclama, mesmo, por mais avaliação autêntica nos JDC.

Só que para a avaliação ser verdadeiramente autêntica e estar alinhada com o ensino, não se devem avaliar separadamente aspectos do desempenho no jogo que são unos. Como as técnicas não são independentes da táctica, mas, antes, instrumentos ao seu serviço (Mahlo, 1969/n. d.), é através da sua utilização adequada que se percebe o nível da compreensão táctica e da capacidade para resolver os problemas tácticos.

Separar a decisão da execução, como faz o GPAI, introduz um artificialismo, que contraria a própria concepção, preconizada pelos seus defensores, de um ensino autêntico dos JDC baseado numa matriz táctica da técnica. A separação só serve para analisar e

perceber as partes depois de compreendido o todo e nunca para apreciar um todo que é, intrinsecamente, uno.

Assim, os procedimentos de avaliação utilizados pelos observadores do grupo experimental 2 foram simples. Se não vejamos:

1 - Utilizaram-se provas de avaliação inspiradas nos objectivos específicos do programa, justamente porque as condições de realização dos objectivos específicos dos PNEF determinam o contexto em que se expressam as competências de um determinado nível, e são, também elas próprias, a referência para se construírem as situações de avaliação, ou seja, a referência para a prova. Deste modo:

- A prova de avaliação que permite distinguir se um jogo e os jogadores são de nível introdução ou de nível elementar no andebol, é a situação de jogo 5x5 num campo reduzido com bola afável;

- A prova de avaliação que permite distinguir se um jogo e os jogadores são de nível elementar ou de nível avançado no andebol, é a situação de jogo 5x5 num campo reduzido;

- A prova de avaliação que permite distinguir se um jogo e os jogadores são de nível avançado ou cumprem nível avançado no andebol, é a situação de jogo 7x7 num campo normal.

2 - Os professores avaliaram o jogo e os jogadores em aproximadamente 8 a 9 minutos, tendo apenas que registar o nível de jogo

3 - Não foi necessário avaliar várias componentes para conseguir-se classificar o desempenho de cada jogador;

4 - Um só avaliador foi suficiente para avaliar o desempenho de todos os jogadores.

A solução foi simples e vai de encontro à tão reclamada necessidade, publicada na literatura sobre avaliação autêntica nos JDC, de se encontrarem procedimentos de avaliação e registo que sejam «amigáveis» para os professores (Harvey, 2006), em alternativa a instrumentos que avaliam situações de jogo autênticas, mas são muito trabalhosos e consomem muito tempo. É, portanto, consistente com o que Allen (2002) preconiza, uma avaliação económica em termos de custos e de tempo dispendido pelos intervenientes.

Na perspectiva de Stobart (2010), é muito difícil garantir que qualidades formais da avaliação, como a validade e a fiabilidade, e os custos dessa mesma avaliação, se conjuguem numa relação aceitável. O autor considera que, por vezes, a avaliação é válida, mas não estando garantidos acordos entre os avaliadores, reduz-se a fiabilidade. No caso do nosso estudo os avaliadores utilizaram os níveis de jogo como referência e, durante o

episódio de instrução, acertaram-se para utilizar, também, os mesmos critérios de avaliação e indicadores de observação.

Outras vezes, quando não são avaliados todos os domínios considerados importantes, a fiabilidade mina a validade, como no exemplo que o mesmo autor dá dos exames em língua inglesa que não avaliam a expressão e a compreensão oral. Também aqui, o nosso estudo contrasta com este exemplo. A fiabilidade não minou a validade, porque as competências prioritárias que estão expressas nos objectivos de cada nível foram avaliadas.

Finalmente, o autor refere que uma avaliação com grande validade e fiabilidade pode implicar recursos avultados. No caso do nosso estudo, os custos foram mínimos, a avaliação não ocupou muito tempo, nem foram utilizados muitos recursos.

### 3.3. Análise das diferenças de classificações produzidas por cada grupo nos dois momentos (conjunto de jogos A e conjunto de jogos B)

Observou-se uma melhoria estatisticamente significativa da soma de erros de avaliação do conjunto de jogos A (jogos 1, 2 e 3) para o conjunto de jogos B (jogos 4, 5 e 6) em qualquer dos três grupos. No grupo de controlo do conjunto de jogos A ( $M = 8.92$ ,  $DP = 4.78$ ) para o conjunto de jogos B ( $M = 7.51$ ,  $DP = 4.20$ ) ( $t(38) = 2.561$ ;  $p = 0.015$ ). No grupo experimental 1 do conjunto de jogos A ( $M = 9.14$ ,  $DP = 4.29$ ) para o conjunto de jogos B ( $M = 4.14$ ,  $DP = 2.47$ ) ( $t(57) = 7.824$ ;  $p < 0.001$ ). No grupo experimental 2 do conjunto de jogos A ( $M = 9.24$ ,  $DP = 4.04$ ) para o conjunto de jogos B ( $M = 2.59$ ,  $DP = 2.29$ ) ( $t(53) = 12.651$ ;  $p < 0.001$ ). A figura 3 ilustra os resultados obtidos.

Tabela 18. Estatística descritiva ( $M$ ,  $DP$ ) da soma de erros por grupo (controlo, experimental 1 e experimental 2) em cada conjunto de jogos (1+2+3 e 4+5+6), e valores do teste  $t$ -Student para amostras emparelhadas relativa à comparação dos erros entre os dois momentos.

Grupo	Jogos		$t$ -Student amostras emparelhadas		
	1+2+3 Erros $M(DP)$	4+5+6 Erros $M(DP)$	$t$	$gl$	$p$
Controlo	8.92 (4.78)	7.51 (4.20)	2.561	38	0.015
Experimental 1	9.14 (4.29)	4.14 (2.47)	7.824	57	0.000
Experimental 2	9.24 (4.04)	2.59 (2.29)	12.651	53	0.000

Os dados desta tabela estão expressos sob a forma de gráfico na Figura 3. que representa os três grupos (controlo, experimental 1 e experimental 2) pelas médias da soma de erros de classificação.

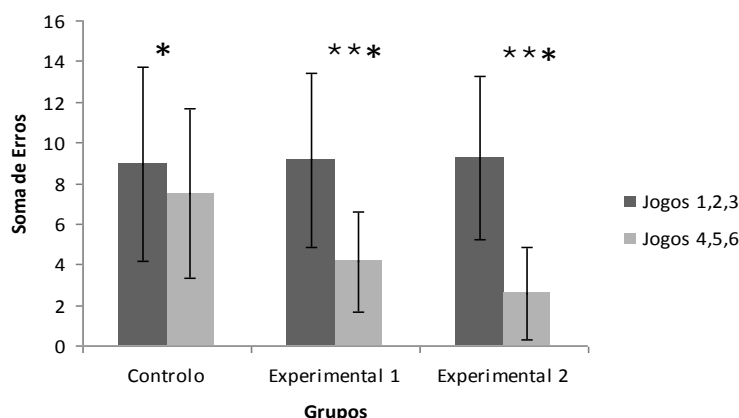


Figura 3. A soma de erros de avaliação média dos observadores nos três grupos (controlo, experimental 1 e experimental 2) no conjunto de jogos A (jogos 1, 2 e 3) vs. no conjunto de jogos B (jogos 4, 5 e 6). A soma de erros de avaliação média no conjunto de jogos B é significativamente menor que no conjunto de jogos A. Grupo de controlo ( $t(38) = 2.561$ ;  $p = 0.015$ ). Grupo experimental 1 ( $t(57) = 7.824$ ;  $p < 0.001$ ). Grupo experimental 2 ( $t(53) = 12.651$ ;  $p < 0.001$ ). As barras de erro representam o desvio padrão.

Para além de se verificarem diferenças estatisticamente significativas entre os grupos relativas à soma de erros de classificação no conjunto de jogos B, verifica-se, também, que todos os grupos (controlo, experimental 1 e experimental 2) tiveram melhorias estatisticamente significativas cometendo menos erros de avaliação no conjunto de jogos B (jogos 4, 5 e 6), os que foram observados depois da intervenção, do que no conjunto de jogos A (jogos 1, 2 e 3) observados antes da intervenção. Destaca-se o resultado de que não só os grupos experimentais melhoraram, como inclusivamente o grupo de controlo também melhorou.

Esta evidência de melhoria das classificações de todos os grupos (controlo, experimental 1 e experimental 2) merece uma reflexão sobre da organização do quadro experimental.

Sendo os conjuntos de jogos A e B equivalentes em termos gerais, i.e. níveis de jogo e dificuldades de classificação, apresentámos aos observadores episódios de jogo distintos interpretados por jogadores diferentes.

Tratando-se de objectos distintos a observar, assegurámos que não se verificaria um efeito de memória do 1º momento para o 2º momento de observação, separados por um intervalo de meia hora, em que se efectuaram os tratamentos (instrução preparatória da observação do 2º momento, conjunto de jogos B).

Estas condições experimentais colocam uma elevada exigência para a probabilidade de ocorrer um efeito do tratamento a que foram submetidos os grupos experimentais 1 e 2 por contraste com o grupo de controlo (tratamento placebo).

Ora, ao verificarmos que não só esse efeito ocorreu para os dois grupos experimentais, mas também para o grupo de controlo, havendo um efeito geral de diminuição de erros de classificação, o que nos leva a inferir que se trata de um efeito directo de aprendizagem do conjunto dos sujeitos no contexto do quadro experimental.

A evidência deste efeito é reforçada pelos resultados que mostram não haver diferenças das classificações produzidas por especialistas comparados com não especialistas e de professores experientes comparados com os estagiários.

Podemos interpretar esta tendência como um efeito de contexto, concretamente um efeito da oportunidade de treinar a observação discriminativa e a classificação de jogadores em situação de jogo.

Por outro lado, este efeito de aprendizagem ocorreu num contexto em que se utilizaram «registos permanentes» (Siedentop, 1991), concretamente, os vídeos elaborados pelos professores da Escola Secundária Dona Luísa de Gusmão a partir das suas provas de avaliação interna e que são utilizados para verificar a fiabilidade das classificações e para o treino de avaliação durante os processos de aferição de critérios. Na análise que Darden (1999) faz da utilização do vídeo como instrumento de apoio ao ensino, enfatiza a importância de integrar esta tecnologia na área da avaliação em EF. Justamente, a utilização do vídeo nesta experiência para treino da avaliação e dos procedimentos de avaliação parece ter sido muito poderosa, já que todos os grupos melhoraram.

Importa, finalmente, relacionar esta melhoria geral das classificações com a evidência de que os grupos experimentais produziram avaliações válidas e fiáveis, na classificação do conjunto de jogos B, diferenciando-se significativamente do grupo de controlo, em especial o grupo experimental 2, que para além disso se distingue significativamente do grupo experimental 1.

Tal relação significa que podemos refutar, no quadro destes resultados, que se circunscreva este método de classificação ao modelo de avaliação pedagógica do Grupo de EF da Escola Secundária de D. Luísa de Gusmão (2003-2012).

Visto que este quadro experimental foca uma parte desse modelo, como referência, parece-nos interessante a exploração de outras hipóteses e soluções que integram o modelo, numa linha de investigação centrada na avaliação autêntica.

## **Conclusões**

Os resultados parecem demonstrar que se podem produzir classificações consistentes a partir de provas de avaliação autêntica nos JDC.

A actividade referente para a aprendizagem e para a avaliação está na dinâmica global, na circunstância total que é o jogo e que representa o valor específico dos JDC, sendo que as provas de avaliação autênticas nos JDC são as que representam, precisamente, essa circunstância total.

Foi possível classificar os jogadores em prova, partindo do jogo, porque os observadores utilizaram uma referência para a avaliação – a definição dos níveis de jogo que está sistematizada nos PNEF.

Os níveis especificados nos PNEF têm sido uma referência para o ensino e para a aprendizagem. Trata-se, agora, de também serem uma referência para a avaliação autêntica, particularmente onde esta avaliação é mais difícil, as matérias interactivas de grupo.

Mas os níveis, para além de permitirem situar o jogador no jogo, também permitem situar o próprio jogo. Ao atribuir significado à actividade representativa de um certo grau de competência prática, está-se a categorizar o jogo, e isso, permite classificá-lo. Pode-se afirmar que o jogo jogado pelo conjunto de jogadores que se observa, é um jogo de nível Introdução, ou de nível Elementar, ou de nível Avançado, ou seja, pode-se classificar a própria prova.

Por outro lado, a utilização de critérios de avaliação e de indicadores de observação, definidos no protocolo de avaliação que do Grupo de Educação Física da Escola Secundária D. Luísa de Gusmão e utilizados neste procedimento experimental, garantiu os avaliadores passassem a situar os jogadores e o jogo com mais facilidade. Os critérios de avaliação, ao representarem os objectivos da aprendizagem mais valorizados em cada nível, ajudaram os observadores a distinguir, não só os níveis de jogo, como também a determinar quando é que o desempenho, conhecimentos e atitudes evidenciados representavam o nível considerado, cumprindo-o. Os indicadores de observação, representando as características específicas identificativas da qualidade que o jogo apresenta num dado nível, revelaram-se uma ajuda preciosa, para determinar se os critérios de avaliação eram cumpridos, ou não.

Foi, também, muito relevante verificar que os observadores com classificações mais consistentes foram aqueles que para além de conhecerem e tomarem os níveis do programa como referência para avaliar, também partiram da classificação do jogo para chegar à classificação individual.

Como o padrão é o do grupo, já que a actividade de grupo é a essência da própria matéria, parece-nos mais correcto partir do grupo para o individual, do todo para a parte, da síntese para a análise, do jogo para o jogador. Além disso, demonstra-se que em termos instrumentais e práticos esta abordagem é fácil e económica.

De facto, o jogo que se observa é representativo da produção colectiva criada por todos os jogadores intervenientes nesse jogo. Esse produto de grupo, que é sempre original e único, resulta de uma complexa conjugação de acções que reflecte, simultaneamente, dinâmicas de cooperação entre colegas de equipa e de oposição entre adversários.

A sistematização em níveis de jogo, distinguindo conteúdos e a forma como o domínio da matéria pode progredir, é representativa, também, de diferentes jogos com qualidades próprias, ou seja, de produções colectivas que têm diferentes qualidades que podem ser observadas.

Ao decidir qual o nível de jogo observado, a tarefa de classificar individualmente os jogadores ficou muito facilitada, porque, precisamente, esse nível de jogo reflectia a qualidade de jogo do conjunto de jogadores observados. Ficava, apenas, por decidir quem se destacava, ou seja, quem é que podia retirar ou acrescentar qualidade a esse produto de grupo, não por ser melhor ou pior que o conjunto dos jogadores, mas por ser de outro nível de jogo diferente do observado.

Por outro lado, pareceu ficar demonstrado que é possível avaliar jogadores, tomando como referência níveis de especificação da matéria, sem haver necessidade de uma análise das características particulares do desempenho individual.

Como já considerámos anteriormente, cada jogador contribui individualmente para a produção colectiva. Esse contributo traduz-se num desempenho em jogo. No entanto, o produto de grupo não resulta, simplesmente, da soma dos desempenhos individuais. A coordenação entre jogadores, as interacções complexas entre colegas de equipa e oposições, as dinâmicas relacionais, e outras, tornam o desempenho colectivo, pelo menos, diferente da simples soma das partes.

O jogo e o desempenho de grupo no jogo são a grande referência de validade. Ao contrário de avaliar cada uma das partes e depois somá-las, como se a avaliação de actividades de grupo fosse simples aritmética, trata-se de em primeiro lugar compreender o todo representado pelo desempenho colectivo no jogo, porque só o todo permite perceber e analisar cada uma das partes. De facto, as qualidades da dinâmica global do jogo perdem-se, se os seus elementos aparecerem decompostos em tarefas discretas.

Mas da mesma forma que o produto colectivo não é a soma de desempenhos individuais, também a avaliação do desempenho individual não é a soma de características particulares, que vistas isoladamente podem fazer perder o essencial da relação entre elas.



Ser coerente com um ensino autêntico é, também na avaliação, não retirar significado ao todo que o jogo representa. Como vimos anteriormente, a tendência dos vários modelos de ensino do jogo, que se constituíram em oposição ao ensino centrado no treino descontextualizado das técnicas, é a do envolvimento dos alunos no jogo real, compreendendo e resolvendo os problemas que ele coloca, com as habilidades técnicas a ser aprendidas em função da lógica do jogo e não o contrário. Assim, neste quadro, para que haja um alinhamento com o ensino, também a avaliação deve partir da circunstância total que é o jogo para o jogador, da referência de grupo, que é o desempenho colectivo, para o indivíduo, e da síntese do desempenho individual globalmente observado, para a análise das componentes que caracterizam esse desempenho.

Observou-se, ainda, que os procedimentos adoptados nos grupos experimentais revelaram ser eficientes. Não foram necessários nem recursos especiais, nem muito tempo.

Em situações de avaliação autêntica, numa matéria especialmente difícil de avaliar, os observadores fizeram apreciações válidas sobre o desempenho dos alunos, sem ocuparem muito tempo, nem utilizarem procedimentos de avaliação e registo trabalhosos.

Este tipo de procedimentos de avaliação ainda oferecem a vantagem de se poder avaliar e ensinar ao mesmo tempo. De facto, como o contexto em que se dá a avaliação pode ser semelhante ao das aulas normais, não perturbando o seu funcionamento, professores que dominem os níveis do programa e que estejam treinados na sua observação e avaliação, conseguem facilmente registar níveis de jogo e, simultaneamente, manterem-se focados na instrução, fornecendo feed-back aos alunos em plena aula.

Os resultados parecem demonstrar que a classificação holística do jogo pode ser válida, mas esta solução simples não é simplista. Conseguir em relativamente pouco tempo avaliar todos os jogadores e o jogo tendo apenas que dar a nota I, E ou A, comporta passos complexos da parte do avaliador. Implica conhecer o jogo em causa e a sua didáctica específica, conhecer os níveis de jogo e os conteúdos que lhes são inerentes, ter a capacidade de observar os comportamentos e a dinâmica global do jogo para poder distinguir os níveis evidenciados e, ainda, ter a intersubjectividade resolvida pela aferição de critérios com os seus pares. É a articulação complexa destes aspectos, em termos da sua competência individual, que permite a solução simples nos procedimentos de avaliação.

#### *Implicações para a avaliação dos alunos*

Quando se avalia nos Jogos Desportivos Colectivos, pode-se situar cada aluno e o próprio jogo em cada um dos três níveis em que se diferencia o tratamento da matéria. Se houver consistência nos dados recolhidos, constitui informação muito relevante saber se um aluno deve aprender o programa de nível introdução ou o de nível elementar, ou saber qual é o nível de jogo em que maioritariamente se situa o conjunto de alunos de uma turma, de

uma escola ou de uma região. Essa informação possibilita que cada professor, o conjunto de professores de um GEF ou de uma escola, ou mesmo a administração central ou regional, possam tomar decisões ao nível da didáctica, do planeamento, do currículo, da implementação de recursos ou da formação de professores. Se não, vejamos:

- Na avaliação formativa, com esses dados, o professor pode organizar as suas opções didácticas, planeando os passos seguintes do processo de aprendizagem.

Em relação ao conjunto da turma, pode mais facilmente escolher as situações em contexto de jogo que melhor servem uma dada etapa de aprendizagem e, também, as situações analíticas que permitem resolver problemas que não conseguiram ser resolvidos em contexto de jogo. Pode definir melhor a composição dos grupos, misturando alunos com níveis de jogo diferentes ou fazendo grupos de nível, consoante a opção pedagógica do momento for a formação de grupos heterogéneos ou homogéneos.

Em relação a cada aluno, pode mais facilmente escolher as situações de treino e aprendizagem que permitam aprender ou aperfeiçoar as competências para cumprir um dado nível. Pode, também, promover a autoavaliação e a auto-regulação da aprendizagem, porque o aluno fica na posse de dados que lhe permitem identificar o que lhe falta para cumprir objectivos de aprendizagem e a distância a que está dos objectivos finais de ano.

- Na avaliação sumativa, essa informação pode ser utilizada para a atribuição de uma classificação em Educação Física.

Conforme consta das Metas de Aprendizagem, que reafirmam as normas de referência para o sucesso em EF, é com base nos níveis cumpridos em categorias diferentes que se sabe se cada aluno atingiu o sucesso. Por exemplo, no 9º ano, na área das actividades físicas, é necessário cumprir 6 níveis introdução e 1 nível elementar para se ter sucesso, sendo obrigatório que dois desses níveis se cumpram em JDC diferentes. Assim, saber qual o nível de jogo de um aluno num determinado JDC tem uma correspondência directa com o cumprimento do nível de sucesso num ano ou ciclo de escolaridade.

- Nas provas de aferição, se os dados forem agregados e convenientemente processados, pode-se ter informação credível para monitorizar as aprendizagens e os progressos reais dos alunos, para caracterizar a população escolar, ou avaliar o sistema de Educação Física Escolar ou, parte dele.

Cabe à inteligência colectiva das equipas pedagógicas, envolvidas na avaliação, a interpretação dos dados recolhidos. Quem está em contacto e estuda a realidade que se quer avaliar, pode interpretá-la e tomar decisões, ou seja agir com base nos resultados.

Nessa abordagem, podemos destacar alguns aspectos fundamentais, a partir da nossa pesquisa.

Os indicadores de observação permitem olhar para as características gerais que o jogo apresenta e identificar um nível. Esta forma de identificar a qualidade do jogo num dado nível pode contribuir para que se consigam dividir os níveis I, E e A. É o que recomendamos. Embora respeitando a sistematização inicial, podemos conseguir que a extensão de conteúdos em cada nível diminua. Tal pode ter implicações no ensino, mas, também, na avaliação. Se considerarmos a existência de partes de nível, isso pode ajudar a melhor situar o jogo e cada jogador no jogo.

De facto, é diferente um jogador de nível introdução que não se desmarca e que não decide bem quando deve progredir em drible ou passar a bola, daquele jogador que decide e faz isto bem, mas não consegue ainda rematar em salto. Em vez de considerarmos que ambos são I, passamos a considerar na codificação que um é I- e o outro I+. Nada disto perverte a sistematização inicial feita pelos autores dos programas e, no caso do GEF da ESDLG, tem-se revelado muito útil na aplicação pedagógica dos programas. Esta questão também se coloca da mesma maneira no nível elementar. É diferente o jogador que apesar de ainda não dominar a marcação de controlo, consegue ultrapassar adversários directos, fintando e rematando em apoio e em suspensão, daquele jogador que ainda não tem essas competências ofensivas. Também aqui, em vez de considerar-se que ambos são E, poder-se-ia considerar que um é E+ e o outro E-.

Há jogos mais difíceis de avaliar por integrarem jogadores e indicadores de dois níveis. São os jogos de «fronteira» entre níveis, aqueles que apresentam características tais, que não permitem uma identificação clara dum determinado nível. Neste tipo de jogos, ou existe um número igual de jogadores a fazer parte de níveis diferentes, e.g. quatro jogadores de nível introdução e quatro jogadores de nível elementar, ou poder-se-á, ainda, estar em presença de vários jogadores de «fronteira».

A ocorrência desta dificuldade tem repercussões diferentes e, também, soluções diferentes dependendo do enquadramento de avaliação em que nos encontremos. Vejamos dois tipos de repercussões e soluções:

No quadro da avaliação formativa a repercussão é a que resulta dessa mesma avaliação. Se estamos perante uma situação de «fronteira» é com esses dados que o professor organiza as suas opções didácticas. Como não há que certificar ou atribuir uma nota, não é necessário nenhuma solução especial, apenas tomar decisões a partir da interpretação dessa realidade que se observou. Por exemplo, trabalhar características ou aspectos que se identificaram como necessários para que um dado nível se cumpra.

No quadro da avaliação sumativa, se o jogo é de «fronteira», isso também não constitui problema. Qualquer que seja a decisão tomada sobre qual é o nível de jogo observado, os jogadores a destacar são, sempre, os do outro nível. Por exemplo, num jogo

que tenha 4 jogadores do nível introdução e 4 jogadores do nível elementar, se os avaliadores decidirem que o jogo é de nível introdução, devem destacar os jogadores de nível elementar, se decidirem que o jogo é de nível elementar, devem destacar os jogadores de nível introdução, o que em termos de certificação ou atribuição de nota terá, exactamente, as mesmas consequências.

Por outro lado, se houver jogadores, eles próprios, individualmente, que são de «fronteira», sugere-se que os avaliadores tomem uma decisão em função de algo que esteja codificado à partida. Confirmada a dúvida, uma solução é classificá-los no nível seguinte. Por exemplo, se o aluno está na «fronteira» do nível introdução com o nível elementar, situá-lo no nível elementar.

Pode também acontecer os avaliadores não conseguirem tomar uma decisão, não por qualquer aspecto relacionado com o desempenho do aluno em causa mas, porque as características dos outros jogadores não o permitem. Duas circunstâncias podem estar na origem desta dificuldade, a saber:

A melhor qualidade dos outros intervenientes esconde ou dissimula a possibilidade desse aluno revelar os comportamentos que se querem observar;

Ou, pelo contrário, sendo o seu desempenho superior ao dos outros, surge uma dúvida sobre qual é o seu verdadeiro nível de jogo, já que essa superioridade pode dever-se, apenas, à debilidade dos outros intervenientes.

Perante dúvidas, tendo havido comportamentos que não puderam ser observados ou completamente esclarecidos porque o contexto de jogo não o permitiu durante o tempo em que decorreu a observação, podem utilizar-se duas possibilidades:

Colocar o aluno em nova situação de jogo, representativa de um nível de jogo diferente, com jogadores que proporcionem um contexto que esclareça o que se quer ver esclarecido;

Ou colocar o aluno numa situação de «análise de pormenor».

As situações de «análise de pormenor» são tarefas contextualizadas, mas mais simples que o jogo total. Por exemplo, se não se observa que um determinado jogador finaliza rematando em salto, porque nunca esteve numa situação de isolado face ao guarda redes, na zona central, e sendo essa competência critério para considerar que um jogador cumpre o nível introdução, podem-se utilizar situações de jogo em superioridade numérica, tipo 2X1, que mantêm alguma contextualização, embora mais simples, e permitem mais oportunidades de finalização.

Não é por um jogador rematar com sucesso na situação de 2X1 que se pode considerar, com absoluta certeza, que ele o consegue fazer na situação de referência, em 5X5, que é mais complexa. Mas, indicia uma forte possibilidade de isso ocorrer.

Na avaliação formativa e na avaliação sumativa interna, não constitui problema esclarecer as dúvidas nas formas que preconizámos, já que, o professor de uma dada turma pode constituir grupos de alunos com características semelhantes, tendencialmente pertencentes a um mesmo nível de jogo. Na turma, o professor pode sempre esclarecer as dúvidas sobre um jogador colocando-o noutra situação de jogo com um contexto previsivelmente mais adequado ao nível desse jogador. Também numa prova de avaliação interna, tipo prova global ou prova de aferição o júri pode adoptar o mesmo procedimento.

Já numa prova de avaliação externa, tipo prova de aferição, se o grupo de alunos que nela participam for constituído por amostra, aumenta a possibilidade de haver alunos de diferentes níveis. No entanto, como este tipo de provas não tem como objectivo certificar cada aluno individualmente, ou seja, não está em causa uma decisão de progressão ou a atribuição de uma classificação individual, então, carece de importância optar por uma das classificações possíveis relativamente aqueles jogadores que provocam dúvidas. Como os dados se referem ao conjunto dos alunos pode bastar que identifiquemos o nível de jogo e, assim, não ser necessário discriminar os níveis de cada jogador.

#### *Implicações na organização das escolas*

Um aspecto relevante a destacar é a constatação de que o tratamento que os grupos experimentais tiveram permitiu uma concertação implícita entre professores, evidenciada na redução das suas divergências de classificação. Em provas de avaliação autênticas que não são repetíveis muitos avaliadores dos grupos experimentais foram mais consistentes nas suas classificações no segundo momento de observação.

Na verdade, o jogo é a grande referência de validade e está na sua autenticidade a garantia de se poderem fazer inferências válidas. É nele que se pode observar o que um aluno sabe e é capaz de fazer. Já a construção do acordo entre os avaliadores quando avaliam o jogo e os jogadores, a fisiologia da coesão do grupo, não é espontânea. A «validação» está resolvida no jogo, mas só o treino da avaliação determina a consistência na avaliação. Garantir que o jogo possa produzir resultados consistentes exige treino dos observadores.

Sendo a fiabilidade uma componente da validade, não nos interessa classificações que apesar de consistentes sejam erradas, ou seja, classificações que falhem ao discriminar a representação do valor de um dado desempenho. Classificações consistentemente erradas conduzem a conclusões que não são válidas.

O esforço deve, então, ser investido no assegurar da consistência da avaliação, sendo necessária formação específica no conhecimento da matéria e formação específica e treino de procedimentos de observação e registo.

Como por hábito a avaliação é vivida em solidão, as tarefas docentes associadas à avaliação são vividas na componente individual e na componente lectiva. É durante o tempo da componente individual que se preparam e elaboram os testes e provas e, posteriormente, se corrigem e classificam, e é nos tempos da componente lectiva que são aplicados aos alunos. A componente não lectiva, pela sua natureza, devia contemplar este tipo de trabalho entre pares.

Ora, mudar a cultura de escola passa por implementar uma cultura de avaliação assente em práticas colectivas. É nas equipas pedagógicas que se resolve a intersubjectividade e se constrói o acordo entre pares. Mas dessas práticas colectivas também deveriam fazer parte, entre outras, o processo de aferição de critérios, a interpretação dos dados recolhidos, particularmente, em provas de avaliação interna, ou em provas de avaliação externa como os exames ou as provas de aferição, e a tomada de decisões que permita agir com base nos resultados.

Precisamos, portanto, de estruturas mais sofisticadas de funcionamento, de mais projecto, autonomia e diferenciação ao nível organizacional para que a avaliação seja verdadeiramente estruturante. Nesta perspectiva, é necessário que as estruturas pedagógicas e de gestão das escolas repensem o tempo atribuído na distribuição do serviço docente, contemplando tempos na componente não lectiva para o trabalho colectivo entre professores. É, também, necessário que estas mudanças sejam facilitadas por lideranças que queiram partilhar responsabilidades e queiram dar força ao trabalho colectivo e às lideranças intermédias que emergem do exercício de boas práticas.

O investimento neste trabalho colectivo seria recompensado. Esta mudança incrementaria fortemente a formação recíproca entre professores, animaria pedagogicamente as escolas, e produziria resultados mais consistentes, e portanto, mais credíveis para alunos, professores, pais e decisores, aumentando a confiança social na escola e contribuindo, em cada uma, para o desenvolvimento e aperfeiçoamento das suas pessoas e dos seus projectos.

Por outro lado, se em Portugal é reconhecida a debilidade da consistência na avaliação interna e se este episódio de instrução dos observadores centrado nos níveis de especificação das matérias garantiu que se produzissem classificações consistentes em matérias tão difíceis de avaliar como são os JDC., então, será pertinente recomendar o seguinte:

1º - Que os Grupos de Educação Física elaborem:

- Protocolos de avaliação que contenham situações de avaliação, inspiradas nas condições de realização dos objectivos específicos dos programas, com critérios de

avaliação que permitam distinguir os diferentes níveis de especificação das matérias, e no caso dos jogos colectivos, com indicadores de observação que ajudem a situar o jogo.

- Processos de treino de avaliação, em que os elementos fundamentais sejam o estudo dos níveis de especificação dos programas e a aferição de critérios centrada na observação de alunos em actividade de jogo, ao vivo ou em vídeo.

- Vídeos representativos dos diferentes níveis (I, E e A) de cada JDC, actualizados regularmente como «registos permanentes», que permitam o treino de avaliação no trabalho colectivo dos agrupamentos de EF.

2º - Que os programas de formação inicial e de formação contínua de professores de Educação Física, em cadeiras de didáctica dos JDC, incluam o estudo dos níveis de especificação dos programas e o treino de observação para avaliar, permitindo que se aprenda a situar o jogo e os jogadores num dado nível do programa.

#### *Futuras investigações*

Se o nível de jogo observado reflecte a qualidade de jogo produzido pelo conjunto de jogadores observados, pode constituir uma pista para futuras investigações, saber quantos alunos das duas equipas determinam essa produção colectiva. Será a maioria, metade ou um mínimo qualificado, por exemplo, 3 em 8? A resposta a esta pergunta poderá ajudar não só na organização das situações de avaliação, como também, na formação de grupos para aprender.

É, também, importante tentar perceber quanto tempo de jogo é, efectivamente, necessário para podermos observar e apreciar o desempenho dos jogadores. Será que podem existir diferenças acentuadas, no tempo necessário à observação, associadas a determinados níveis de jogo ou a determinado JDC? Em média, em cada JDC, quanto tempo demoram os professores treinados a avaliar o jogo e todos os seus intervenientes? Saber as respostas facilitará a organização de situações de avaliação, havendo uma melhor previsão da sua duração, particularmente, em provas de avaliação em EF.

Por outro lado, na aplicação dos níveis à classificação dos jogos, se houver muitos «jogadores de fronteira», o nível pode não ser discriminativo, ou seja, não estar bem definido. Um interessante processo de validação dos níveis pode passar por perceber que percentagem de «jogadores de fronteira» é aceitável para considerar que os níveis são discriminativos. Podiam ser feitos estudos, utilizando professores com experiência consistente de avaliação autêntica em EF, para se perceber quais as percentagens mais frequentes de «jogadores de fronteira» que a actual sistematização de níveis provoca.

É, também, importante pesquisar se as dúvidas de apreciação sobre determinados jogadores dizem, apenas, respeito ao facto do seu desempenho não ser absolutamente claro e identificador de um nível ou se, pelo contrário, é o próprio jogo, objectivamente muito

variado e dinâmico, que faz colocar as dúvidas. O jogo não se repete e pode não proporcionar a oportunidade de observar os comportamentos típicos que evidenciam um determinado nível.

Como os jogos são todos diferentes, mesmo que sempre jogados pelos mesmos jogadores, uma hipótese de investigação é, precisamente, tentar perceber quantos jogos são necessários, e as características que devem ter para esclarecer dúvidas. Quantos jogos são necessários para classificar todos os alunos, sabendo que nem todos os alunos têm oportunidade de expressar os critérios e de claramente mostrar um padrão de desempenho.

No entanto, é necessário cuidado, há casos em que a avaliação não «representa aprendizagem», mas sim, «provoca aprendizagem». É o caso de provas repetidas ou trabalhos de grupo na aula. É importante saber quantos jogos são necessários para provocar aprendizagem e que características devem esses jogos ter, nomeadamente, ao nível da composição entre jogadores? Por exemplo, repetir alguns jogos, onde se misturam alunos do nível elementar com alunos que estão quase a cumprir o nível introdução «I+», pode não ter o mesmo efeito de aprendizagem do que repetir jogos com alunos, todos do nível introdução.

Outra questão pertinente prende-se com o facto de apesar dos indicadores de observação poderem ser mais representativos do todo, do que as estatísticas de jogo, isso não excluir a possibilidade de se usarem dados quantitativos que melhor os ilustrem. Tal, pode até configurar hipóteses de estudo que permitam concluir, por exemplo, quantas situações de finalização, não protagonizadas, devem existir em dez minutos de um jogo de andebol 5X5, para se poder afirmar que esse jogo cumpre o nível introdução, ou quantas vezes a bola deve passar a rede, num jogo de cooperação 2+2 no voleibol, para se poder afirmar que o jogo cumpre o nível introdução.

Sugere-se, também, que em futuras réplicas deste estudo, nos JDC ou em actividades onde a produção colectiva de um grupo pode ser observada, os indicadores de observação sejam apenas objecto de formação no grupo experimental 2. De facto, os indicadores de observação estão fortemente associados a quem em primeiro lugar classifica o jogo. São eles que permitem, isso mesmo, classificar o jogo. Tal procedimento pode permitir com mais clareza distinguir quem conhece os níveis de jogo e os usa na avaliação, daqueles que para além de conhecerem os níveis de jogo e os usarem na avaliação, também utilizam, em primeiro lugar, a classificação do jogo como procedimento de avaliação.

Para além dos Jogos Desportivos Colectivos de invasão e de outros JDC como o Voleibol poder-se-ia estudar a aplicação deste estudo em outras matérias que existem na extensão do currículo da Educação Física. Recomendamos, por isso, que se realizem



estudos em matérias de grupo como as Danças Tradicionais e a Ginástica Acrobática e, ainda, em matérias interactivas que não são de grupo mas que são em pares como as Danças Sociais, os Desportos de Raquetas e os Desportos de Combate.

Finalmente, gostaríamos de referir que, se os conteúdos de uma dada matéria/disciplina forem estruturados por níveis em progressão, traduzindo competências representativas de cada nível de interpretação dessa matéria, então, os procedimentos de avaliação que se utilizaram neste estudo podem ser usados em quaisquer situações de avaliação autêntica de desempenhos de grupo que tenham valor de autonomia e originalidade. Naturalmente, tal necessita de mais desenvolvimento e de maior generalização do campo teórico.

Esta perspectiva pode constituir uma pista para futuras investigações, não só nas matérias de grupo na Educação Física, mas também, em outras áreas do currículo que envolvam situações de desempenho de grupo. É o caso das Línguas Estrangeiras ou do Português, quando em situações de debate entre alunos se pretende avaliar a compreensão e a expressão oral. Esta situação em debate, se for adequadamente estudada, pode ser aproveitada como situação formal de avaliação constituindo-se como alternativa à tradicional prova oral individual porque, obviamente, oferece a vantagem de consumir muito menos tempo. Mas também é o caso do Teatro, quando se quer avaliar uma performance em grupo, da Educação Musical quando se quer avaliar o desempenho colectivo e individual numa Banda ou no Canto Coral, ou de disciplinas como a Matemática, Ciências e outras quando se quer avaliar o trabalho de grupo no âmbito da aprendizagem cooperativa.

#### *Conclusão geral*

Este estudo revelou que é possível produzir classificações consistentes, a partir de provas de avaliação autêntica nos JDC.

Nos JDC as competências, ou as combinações de conhecimentos, habilidades e atitudes, expressam-se no jogo que é a sua actividade cultural referente. Ora, numa matéria da EF, da subárea dos JDC, que é muito difícil de avaliar porque o desempenho é colectivo, estando cada jogador dependente dos colegas de equipa, dos oponentes e da situação, e o contexto de jogo é caracterizado pela abertura, incerteza, imprevisibilidade, sendo improvável que as situações se repitam, foi possível utilizar uma avaliação holística caracterizada por procedimentos de avaliação simples, que não são trabalhosos nem consomem muito tempo.

Como vimos, a constatação desta possibilidade coloca, também, a hipótese de se conseguirem avaliações autênticas consistentes, com custos aceitáveis, em outras matérias do currículo para além das matérias de grupo na EF.

Trata-se de utilizar procedimentos de avaliação semelhantes aos dos observadores que, nesta experiência, partiram do jogo e do desempenho de grupo para, assim, melhor conseguirem classificar o desempenho individual. São, então, necessários:

- Níveis de matérias, suficientemente discriminativos dos conteúdos e especificados em progressão;
- Critérios de avaliação que permitam fazer a distinção entre esses níveis;
- Indicadores de observação que ajudem a situarem o desempenho colectivo e individual.

No entanto, só com treino de avaliação e aferição de critérios se garante a consistência na avaliação. É através da aprendizagem proporcionada pelo treino de procedimentos de observação e registo que se melhora a capacidade de avaliar, e é através da aferição de critérios que se reduzem as divergências e a subjectividade, harmonizando as classificações entre os professores.

Mas, para que esta construção do acordo entre pares aconteça são necessárias, outra organização e outra cultura nas escolas. É necessário que as direcções das escolas reconheçam a necessidade do trabalho colaborativo em avaliação e o valorizem, atribuindo-lhe tempo na distribuição do serviço docente. É, também, necessário aprofundar e dar força às boas práticas para que possam ser estudadas e, em seguida, contextualizadamente generalizadas. Assente na colaboração e reflexão docentes, a avaliação autêntica pode ser, de facto, estruturante numa nova cultura de escola.

## Bibliografia

- Afonso, N., & Costa, E. (2009). A influência do PISA na decisão política em Portugal: o caso das políticas educativas do XVII Governo Constitucional Português. *Sísifo/Revista de Ciências da Educação*, 10, 53-63. Acedido em 15 de Julho de 2012 em <http://sisifo.fpce.ul.pt/>
- Allal, L. (2001). La pédagogie par objectifs est-elle compatible avec des situations d'apprentissage complexe? In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 139-140). Porto Alegre: Artmed.
- Allal, L. (2010). Assessment and the Regulation of Learning. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (pp. 348-352). Oxford: Elsevier.
- Allen, R. (2002). Using Assessment Data to Monitor Physical Education Programs. *The Journal of Physical Education, Recreation & Dance*, 73(8), 25-31.
- Almond, L. (1986a). Primary and secondary rules in games. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 73-74). Loughborough: University of Technology.
- Almond, L. (1986b). Reflecting on themes: A games classification. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 71-72). Loughborough: University of Technology.
- Antunes, S., & Sampaio, J. (2003). O ensino do Basquetebol através do conhecimento táctico. In S. J. Ibáñez & M. M. Macias (Eds.), *Propuestas para la mejora en el proceso de formación y rendimiento en baloncesto*. Acedido em 15 de Julho de 2012 em <http://www.planetabasket.pt/dev/images/pdf/>
- Araújo, D. (2002). A auto-organização da acção táctica: Comentário a Costa, Garganta, Fonseca e Botelho (2002). *Revista Portuguesa de Ciências do Desporto*, 3(3), 87-93.
- Araújo, D., & Davids, K. (2009). Ecological approaches to cognition and action in sport and exercise: Ask not only what you do, but where you do it. *International Journal of Sport Psychology*, 40(1), 5-37.
- Barret, S. (2001). The impact of training on rater variability. *International Education Journal*, 2(1), 49-58. Acedido em 15 de Julho de 2012 em <http://ehlt.flinders.edu.au/education/iej/articles/v2n1/barrett/barrett.pdf>
- Barreto, H., Beja, E., Bom, L., Carreiro da Costa, F., Damas, C., Lima, T., et al. (1980). *Da actividade lúdica à formação desportiva: Seminário de Metodologia do Basquetebol*. Cruz Quebrada: ISEF.

- Barroso, J. (2005). *Políticas Educativas e Organização Escolar*. Lisboa: Universidade Aberta.
- Barton, P. E. (2008). The Right Way to Measure Growth. *Educational Leadership*, 65(4), 70-73.
- Bayer, C. (1985). Pour une pratique transferable dans l'enseignement des sports collectifs. In CONI (Eds.). *Proceedings of International Congress of Teaching Team Sports* (pp. 198-208); Roma, 1983. Roma: Scuola dello Sport.
- Bom, L. (1980). Da Actividade Lúdica à Formação Desportiva. In H. Barreto, E. Beja, L. Bom, F. Carreiro da Costa, C. Damas, T. Lima, et al. (Eds.), *Da Actividade Lúdica à Formação Desportiva: Seminário de Metodologia do Basquetebol* (pp. 19-44). Cruz Quebrada: ISEF.
- Bom, L., Carvalho, L., Mira, J., Pedreira, M., Carreiro da Costa, F., Cruz, S., Jacinto, J., & Rocha, L. (1989). *Programas Nacionais de Educação Física, 1º – 12º ano*. Projecto de Programas de EF para os Ensinos Básico e Secundário instituído pelo Dec-Lei nº 286/89, de 29 de Agosto. Edição: DGBS-Ministério da Educação.
- Bom, L., Carvalho, L., Mira, J., Pedreira, M., Carreiro da Costa, F., Cruz, S., Jacinto, J., & Rocha, L. (1990). A elaboração do projecto de programas de educação física. *Revista Horizonte*, 6(35), Dossier.
- Bonniol, J. J. (2001a). Entre les deux logiques de l'évaluation, rupture ou continuité? In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 143-147). Porto Alegre: Artmed.
- Bonniol, J. J. (2001b). Recherches et formation, pour une théorie de l'évaluation formative. In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 279-280). Porto Alegre: Artmed.
- Bonniol, J. J., & Vial, M. (2001). *Modelos de Avaliação: Textos fundamentais*. Porto Alegre: Artmed.
- Bracken, B. (2002). Team Skills and Individual Assessment in Team Settings. Beyond Boundaries. *Proceedings of the NCIIA 6th Annual Meeting* (pp. 21-25); Washington, D.C.
- Brooker, R., Kirk, D., Braiuka, S., & Bansgrove, A. (2000). Implementing a game sense approach to teaching high school basketball in a naturalistic setting. *European Physical Education Review*, 6(1), 7-26.
- Brooks, C. M., & Ammons, J. L. (2003). Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessment. *Journal of Education for Business*, 78(5), 268-272.

- Bunker, D., & Thorpe, R. (1982). A model for the teaching of games in secondary schools. *Bulletin of Physical Education*, 18, 7-10.
- Bunker, D., & Thorpe, R. (1986a). From theory to practice. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 11-16). Loughborough: University of Technology.
- Bunker, D., & Thorpe, R. (1986b). Is there a need to reflect on our games teaching. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 25-34). Loughborough: University of Technology.
- Bunker, D., & Thorpe, R. (1986c). The curriculum model. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 7-10). Loughborough: University of Technology.
- Caldwell, B. (2000). A 'Public Good' Test to Transformation of Public Education. *Journal of Educational Change*, 1(4), 307-329.
- Cardinet, J. (2001a). Évaluation interne, externe ou négociée? In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 54-56). Porto Alegre: Artmed.
- Cardinet, J. (2001b). Pour apprécier le travail des élèves. In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 115-117; 121-123). Porto Alegre: Artmed.
- Carreiro da Costa, F., & Piéron, M. (1997). Teaching the curriculum: policy and practice in Portugal and Belgium. *The Curriculum Journal*, 8(2), 231-247
- Carvalho, L. M. (1994). Avaliação das Aprendizagens em Educação Física. *Boletim da Sociedade Portuguesa de Educação Física*, 10/11, 135-151.
- Chow, J. Y., Davids, K., Button, C., Shuttleworth, R., Renshaw, I., & Araújo, D. (2007). The role of Nonlinear pedagogy in physical education. *Review of Educational Research*, 77(3), 251-278.
- Chow, J. Y., Davids, K., Hristovski, R., Araújo, D., & Passos, P. (2011). Nonlinear pedagogy: Learning design for self-organizing neurobiological systems. *New Ideas in Psychology*, 29(2), 189-200.
- Clímaco, M. C. (2005). *Avaliação de Sistemas em Educação*. Lisboa: Universidade Aberta.
- Cohen, E. G., Lotan, R. A., Abram, P. L., Scarloss, B. A., & Schultz, S. E. (2002). Can groups learn? *Teachers College Record*, 104(6), 1045-1068.
- Comédias, J. (2004). Avaliação aferida – Uma necessidade do sistema de Educação Física na escola. *Revista Horizonte*, 19(112), 9-12
- Comédias, J. (2005a). Como organizar as provas aferidas em Educação Física – Proposta de modelo global. *Revista Horizonte*, 20(118), 3-9

- Comédias, J. (2005b). Provas de avaliação aferida em educação física. *Boletim da Sociedade Portuguesa de Educação Física*, 30/31, 177-196
- Comédias, J. (2006). Como organizar as provas de avaliação aferida dos JDC? *Revista Horizonte*, 21(122), 3-7
- Conceição, J. (1994). Avaliação Aferida – para quê? (II). In IIE (Ed.), *Pensar avaliação, melhorar aprendizagem* (Folha F/2). Lisboa: IIE.
- Conselho Nacional das Associações de Professores e Profissionais de Educação Física. (2006). Educação Física Curricular e Desporto Escolar. Moção apresentada no 7º Congresso Nacional de Educação Física. In Petrucci, R., Bom, L., & Mira, J. (Eds.), *CNAPEF – 20 anos* (pp. 165-166). Lisboa: CML.
- Coryn, C. L. S. (2007). The holy trinity of methodological rigor: A sceptical view. *Journal of MultiDisciplinary Evaluation*, 4(7), 26-30.
- Coutinho, C. (2011). *Metodologia de Investigação em Ciências Sociais e Humanas: Teoria e Prática*. Coimbra: Almedina.
- Crahay, M. (2002). *Poderá a escola ser justa e eficaz?* Lisboa: Instituto Piaget.
- Crespo, J. (1979). Teoria e metodologia dos jogos desportivos colectivos. *Ludens*, 3 (2/3), 63-75.
- Darden, G. F. (1999). Videotape feedback for student learning and performance: A learning-stages approach. *The Journal of Physical Education, Recreation & Dance*, 70(9), 40-44.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047-1085.
- De Ketele, J. M., & Gerard, F. M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et Évaluation en Éducation*, 28(3), 1-26.
- De Landsheere, G. (1997). *A pilotagem dos sistemas de educação*. Porto: Edições ASA.
- De Landsheere, V. (1988). *Faire réussir faire échouer: La compétence minimale et son évaluation*. Paris: Presses Universitaires de France.
- Dejong, G., Kokinakis, C. L., & Kuntzleman, C. (2002). The role of assessment in meeting the NASPE physical education content standards. *The Journal of Physical Education, Recreation & Dance*, 73(7), 22-25.
- Doering, N. (2000). Measuring Student Understanding with a Videotape Performance Assessment. *The Journal of Physical Education, Recreation & Dance*, 71(7), 47-52.
- Doolittle, S. (1986). Reflecting on an innovation. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 53-56). Loughborough: University of Technology.

- Dyson, B. Griffin, L. L., & Hastie, P. (2004). Sport Education, Tactical Games, and Cooperative Learning: Theoretical and pedagogical considerations. *Quest*, 56, 226-240. Acedido em 15 de Julho de 2012 em <http://www.sirc.ca/>
- Eckes, T. (2005). Examining rater effects in TestDaf writing and speaking performance assessment: A many-facet Rasch analysis. *Language Assessment Quarterly*, 2(3), 197-221. Acedido em 15 de Julho de 2012 em <http://www.testdaf.de/>
- Eisner, E. (1999). The uses and limits of performance assessment. *Phi Delta Kappan*, 80, 658-660.
- Elias, N. (1997). *Envolvimento e distanciamento*. Lisboa: Publicações Dom Quixote.
- Fay, T., & Doolittle, S. (2002). Agents for change: From standards to assessment to accountability in physical education. *The Journal of Physical Education, Recreation & Dance*, 73(6), 29-33.
- Fernandes, D. (1994). Contornos de uma experiência de avaliação desenvolvida no Instituto de Inovação Educacional (1990-1993). *Boletim da Sociedade Portuguesa de Educação Física*, 10/11, 7-32.
- Fernandes, D. (2005). *Avaliação das Aprendizagens: Desafios às Teorias, Práticas e Políticas*. Lisboa: Texto Editores.
- Fernandes, D. (2006). Para uma teoria da avaliação formativa. *Revista Portuguesa de Educação*, 19(2), 21-50.
- Fernandes, D. (2007). A avaliação das aprendizagens no sistema educativo português. *Educação e Pesquisa*, São Paulo, 33(3), 581-600.
- Fernandes, D. (2008). Para uma teoria da avaliação no domínio das aprendizagens. *Estudos em Avaliação Educacional*, 19(41), 347-371.
- Fernandes, D. (2009). Avaliação das aprendizagens em Portugal: investigação e teoria da actividade. *Sísifo. Revista de Ciências da Educação*, 9, 87-100. Acedido em 15 de Julho de 2012 em <http://sisifo.fpce.ul.pt/>
- Fernandes, D., Neves, A., Campos, C., & Laland, J. (1996). *Das concepções, práticas e organização da avaliação das aprendizagens à formação de professores*. (Relatório do 1º ano do Projecto PI/12/94 financiado pelo Instituto de Inovação Educacional). Manuscrito não publicado, Universidade de Aveiro, Aveiro.
- Fleiss, J. (1981). *Statistical methods for rates and proportion* (2ª ed.). New York: Wiley.
- Fragata, J., & Martins, L. (2004). *O Erro em Medicina: perspectivas do indivíduo, da organização e da sociedade*. Coimbra: Almedina.
- Gallagher, C. W., & Ratzlaff, S. (2008). The road less travelled. *Educational Leadership*, 65(4), 48-53.

- Garganta, J. (1998). Para uma teoria dos jogos desportivos colectivos. In A. Graça & J. Oliveira (Eds.), *O ensino dos jogos desportivos* (3ª ed) (pp. 11-26). Porto: FCDEF.
- Garganta, J. (2009). Trends of tactical performance analysis in team sports: bridging the gap between research, training and competition. *Revista Portuguesa de Ciências do Desporto*, 9(1), 81-89.
- Genthon, M. (2001a). Construire des situations d'apprentissage. In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 201-202). Porto Alegre: Artmed.
- Genthon, M. (2001b). Places et fonctions de l'auto-évaluation dans un dispositif d'apprentissage. In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 246-248). Porto Alegre: Artmed.
- Gerard, F. M. (2006). Les outils ouverts d'évaluation, ou la nécessité de clés de fermeture. In G. Figari & L. Mottier Lopez (Eds.), *Recherche sur l'évaluation en éducation – Problématiques, méthodologies et épistémologie* (pp. 147-154). Paris: L'Harmattan.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.
- Goodson, I. (1997). *A construção social do currículo*. Lisboa: Educa.
- Graça, A. (1998). Os comos e os quandos no ensino dos jogos. In A. Graça & J. Oliveira (Eds.), *O ensino dos jogos desportivos* (3ª ed) (pp. 27-34). Porto: FCDEF-UP.
- Graça, A. (2003). Para uma renovação das concepções de ensino dos jogos desportivos na escola. In A. Prista, A. Marques, A. Madeira, & S. Saranga (Eds.), *Actividade Física e Desporto: Fundamentos e Contextos* (pp. 177-188). Porto; Maputo: Faculdade de Ciências de Educação Física e Desporto, Universidade Pedagógica de Moçambique; Faculdade de Ciências do Desporto e de Educação Física, Universidade do Porto; Faculdade de Medicina, Universidade Eduardo Mondlane, Moçambique.
- Graça, A., & Mesquita, I. (2007). A investigação sobre os modelos de ensino dos jogos desportivos. *Revista Portuguesa de Ciências do Desporto*, 7(3), 401-421.
- Graça, A., Musch, E., Mertens, B., Timmers, E., Mertens, T., Taborsky, F., et al. (2003). *O modelo de competência nos jogos de invasão: Proposta metodológica para o ensino e aprendizagem dos jogos desportivos*. Comunicação apresentada no II Congreso Ibérico de Baloncesto, Cáceres, Spain.
- Greene, J. P., Winters, M. A., & Forster, G. (2004). Testing high-stakes tests: Can we believe the results of accountability tests? *Teachers College Record*, 106(6), 1124-1144.
- Gréhaigne, J. F., & Godbout, P. (1998). Formative assessment in team sports with a tactical approach. *The Journal of Physical Education, Recreation and Dance*, 69(1), 46-51.



- Gréhaigne, J. F., Godbout, P., & Bouthier, D. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education*, 16, 500-516.
- Gréhaigne, J. F. Godbout, P., & Bouthier, D. (2001). The teaching and learning of decision making in team sports. *Quest*, 53, 59-76.
- Gréhaigne, J. F., Caty, D., & Godbout, P. (2010). Modelling ball circulation in invasion team sports: a way to promote learning games through understanding. *Physical Education and Sport Pedagogy*, 15(3), 257-270.
- Gréhaigne, J. F., Richard, J. F., & Griffin, L. L. (2005). *Teaching and learning team sports and games*. New York: RoutledgeFalmer.
- Griffin, L. L., Mitchell, S. A., & Oslin, J. L. (1997). *Teaching sport concepts and skills: A tactical games approach*. Champaign, IL: Human Kinetics.
- Grupo de Educação Física da Escola Secundária Dona Luísa de Gusmão. (2010). *Protocolo de Avaliação em Educação Física*. Manuscrito não publicado, ESDLG, Lisboa.
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research & Development*, 52(3), 67-86. Acedido em 15 de Julho de 2012 em <http://igitur-archive.library.uu.nl/>
- Guskey, T. (2005). *Multiple sources of evidence: An analysis of stakeholders' perceptions of different indicators of student learning*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.
- Hansen, E. J., & Stephens, J. A. (2000). The ethics of learner – centered education. *Change*, 33(5), 41-47.
- Hardman, K. (2008). Physical Education in Schools: A Global Perspective. *Kinesiology*, 40(1), 5-21.
- Harlen, W. (2004). *Can assessment by teachers be a dependable option for summative purposes?* Paper presented at the General Teaching Council for England conference “New Relationships: Teaching, Learning and Accountability”, London, England.
- Harlen, W. (2006). On the Relationship between Assessment for Formative and Summative Purposes. In J. Gardner (Ed.), *Assessment and Learning* (pp. 103-117). London: Sage.
- Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation*, 33, 15-28.
- Harlen, W., & Gardner, J. (2010). Assessment to support learning. In J. Gardner, W. Harlen, L. Hayward, G. Stobart & M. Montgomery (Eds.), *Developing Teacher Assessment* (pp. 15-28). Berkshire: McGraw – Hill.

- Harvey, S. (2004). Teaching Games for Understanding: A study of U19 college soccer players improvement in game performance using the Game Performance Assessment Instrument. In R. Light, K. Swabey, & R. Brooker (Eds.). *Proceedings of the 2<sup>nd</sup> International Conference: Teaching Sport and Physical Education for Understanding* (pp. 11-25); Melbourne, Australia, 11 – 14 December, 2003. Melbourne, Australia: University of Melbourne.
- Harvey, S. (2006). *Effects of Teaching Games for Understanding on Game Performance and Understanding in Middle School Physical Education*. (Doctoral dissertation). Acedido em 15 de Julho, 2012 em <http://scholarsarchive.library.oregonstate.edu/>
- Harvey, S. (2007). Using a generic invasion game for assessment: Conducting authentic assessments for invasion-game knowledge and skills just became easier! *The Journal of Physical Education, Recreation and Dance*, 78(4), 19-25, 48-50.
- Hay, P. J. (2006). Assessment for learning in physical education. In D. Kirk, D. Macdonald & M. O'Sullivan (Eds.), *The Handbook of Physical Education* (pp. 312-325). London: SAGE Publications.
- Herrington, A., & Herrington, J. (2006). What is an Authentic Learning Environment?. In Herrington, T., & Herrington, J. (Eds.), *Authentic Learning Environments in Higher Education*. (pp. 1-14). Hershey PA: Idea Group.
- Hicks, J. (2007). Students views of cooperative learning and group testing. *Radiologic Technology*, 78(4), 275-283.
- Holt, N. L., Streat, L., & Bengoechea, W. B. (2002). Expanding the teaching games for understanding model: New avenues for future research and practice. *Journal of Teaching in Physical Education*, 21, 162-176.
- Hopper, T. (2002). Teaching games for understanding: The importance of student emphasis over content emphasis. *The Journal of Physical Education, Recreation and Dance*, 73(7), 44-48.
- Hopper, T., & Bell, R. (2001). Games classification system: Teaching strategic understanding and tactical awareness. *CAHPERD*, 66(4), 14-19.
- Hubball, H. (2004). Problem-Based Learning Enhancing Games for Understanding in a Youth Soccer Academy Program. In R. Light, K. Swabey, R. Brooker (Eds.). *Proceedings of the 2<sup>nd</sup> International Conference: Teaching Sport and Physical Education for Understanding* (pp. 34-43); Melbourne, Australia, 11 – 14 December, 2003. Melbourne, Australia: University of Melbourne.
- Jacinto, J., Carvalho, L., Comédias, J., & Mira, J. (2001). *Programas Nacionais de Educação Física – Reajustamento*. Revisão dos PNEF (1989) homologada pelo Dec-Lei nº 6

- /2001 e pelo Dec-Lei nº 7 de 2001, de 16 de Janeiro. Edição online da DGIDC-ME. Acedido em 15 de Julho, 2012 em [www.dgfdc.min-edu.pt](http://www.dgfdc.min-edu.pt).
- Júlio, L. (2006). *Abordagem Centrada nos Constrangimentos para o Ensino, Treino e Aprendizagem da Tomada de Decisão no Jogo de Futsal: Proposta integradora das perspectivas da psicologia cognitiva e ecológica para aplicação no contexto escolar*. Comunicação apresentada no Centro de Estudos de Fátima, Fátima.
- Kellaghan, T., & Madaus, G. (2003). External (public) examinations. In T. Kellaghan & D. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 577-602). Dordrecht: Kluwer.
- King, P. E., & Behnke, R. R. (2005). Problems associated with valuating student performance in groups. *College Teaching*, 53(2), 57-61.
- Kirk, D., & McPhail, A. (2002). Teaching games for understanding and situated learning: Re-thinking the Bunker-Thorp model. *Journal of Teaching in Physical Education*, 21, 177-192.
- Kirkendall, D., Gruber, J., & Johnson, R. (1987). *Measurement and Evaluation for Physical Educators* (2ª ed.). Champaign, IL: Human Kinetics.
- Konzag, I. (1991). A formação técnico-táctica nos Jogos Desportivos Colectivos. *Treino Desportivo, II sér.*, 19, 27-37.
- Kupermintz, H. (2004). On the reliability of categorically scored examinations. *Journal of Educational Measurement*, 41(3), 193-204.
- Lames, M., & McGarry, T. (2007). On the search for reliable performance indicators in game sports. *International Journal of Performance Analysis in Sport*, 7(1), 62-79.
- Mahlo, F. (n. d.). *O acto táctico no jogo*. Lisboa: Compendium. (Obra original publicada em 1969)
- Mandigo, J., Butler, J., & Hopper, T. (2007). What is Teaching Games for Understanding? A Canadian Perspective. *Physical and Health Education*, 73(2), 14-20.
- Mandigo, J., & Holt, N. (2002). *The Inclusion of Optimal Challenge in Teaching Games for Understanding*. USA: Department of Physical Education, Brock University.
- Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. A paper in the series: Implementing the state accountability system requirements under the no child left behind act of 2001. Washington, DC: Council of Chief State School Officers.
- Marôco, J. (2011). *Análise Estatística com o SPSS Statistics* (5ª ed.). Pêro Pinheiro: Report Number.

- Martin, A., & Gaskin, C. (2004). An integrated approach to coaching athletes. In R. Light, K. Swabey, R. Brooker (Eds.). *Proceedings of the 2<sup>nd</sup> International Conference: Teaching Sport and Physical Education for Understanding* (pp. 62-69); Melbourne, Australia, 11 – 14 December, 2003. Melbourne, Australia: University of Melbourne.
- Martin, J. M., Kulinna, P. H., & Cothran, D. (2002). Motivating students through assessment. *The Journal of Physical Education, Recreation & Dance*, 73(8), 18-30.
- McCaughtry, N., Sofo, S., Rovegno, I., & Curtner-Smith, M. (2004). Learning to teach Sport Education: Misunderstandings, pedagogical difficulties, and resistance. *European Physical Education Review*, 10(2), 135-155.
- McMillan, J. H. (2005). *The impact of high-stakes test results on teacher instructional and classroom assessment practices*. Virginia: Metropolitan Educational Research Consortium, Virginia Commonwealth University. Acedido em 27 de Julho, 2011 em [www.eric.ed.gov](http://www.eric.ed.gov).
- McTighe, J., & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership*, 63(3), 10-17.
- Meisels, S. J., Liaw, F., Dorfman, A., & Nelson, R. F. (1995). The work sampling system: Reliability and validity of a performance assessment for young children. *Early Childhood Research Quarterly*, 10, 277-296.
- Memmert, D., & Harvey, S. (2008). The Game Performance Assessment Instrument (GPAI): Some concerns and solutions for further development. *Journal of Teaching in Physical Education*, 27, 220-240.
- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers knowledge & application of classroom assessment concepts: Development of the Assessment Literacy Inventory*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Mesquita, I. (2000). Modelação do Treino das Habilidades Técnicas nos Jogos Desportivos. In J. Garganta (Ed.), *Horizontes e Órbitas no Treino dos Jogos Desportivos* (pp. 63-73). Porto: FCDEF-UP.
- Mitchell, S. A., Oslin, J. L., & Griffin, L. L. (2003). *Sport foundations for elementary physical education: A tactical games approach*. Champaign, IL: Human Kinetics.
- Moon, T. R., Brighton, C. M., Callahan, C. M., & Robinson, A. (2005). Development of Authentic Assessment for the Middle School Classroom. *Journal of Secondary Gifted Education*, 16, 119-133. Acedido em 19 Outubro, 2007 em <http://www.questia.com/>
- Nadeau, L., Godbout, P., & Richard, J. F. (2008). Assessment of ice hockey performance in real-game conditions. *European Journal of Sport Science*, 8(6), 379-388.

- National Association of State Boards of Education. (1999). Setting Cut Scores on Large-Scale Assessment. *The State Board Connection: Issues in Brief*, 19, 2, 1-12.
- Nazario, L. (2005). *The use of alternative assessments in physical education: Why some do but many more don't*. (Doctoral dissertation). Acedido em 27 de Julho, 2011 em <http://etd.lib.fsu.edu/theses/>.
- Nevett, M., Rovegno, I., Babiarz, M., & McCaughtry, N. (2001). Changes in basic tactics and motor skills in an invasion-type game after a 12-lesson unit of instruction. *Journal of Teaching in Physical Education*, 20(4), 352-369.
- Newell, K. M. (1986). Constraints on the development of coordination. In M. Wade & H. T. A. Whiting (Eds.), *Motor Development in Children: Aspects of Coordination and Control* (pp. 341-360). Dordrecht: Martinus Nijhoff.
- Noizet, G., & Caverni, J. P. (1985). *Psicologia da avaliação*. Coimbra: Coimbra Editora.
- Organisation for Economic Co-operation and Development. (2008). *Assessment for learning – The case for formative assessment*. Paper presented at the OECD/CERI International Conference “Learning in the 21<sup>st</sup> Century: Research, Innovation and Policy”, Paris, France.
- Oslin, J. L., Mitchell, S. A., & Griffin, L. L. (1998). The game performance assessment instrument (GPAI): Development and preliminary validation. *Journal of Teaching in Physical Education*, 17(2), 231-243.
- Pacheco, J. A. (1994). *A avaliação dos alunos na perspectiva da reforma*. Porto: Porto Editora.
- Passos, P., Batalau, R., & Gonçalves, P. (2006). Comparação entre as abordagens ecológica e cognitivista para o treino da tomada de decisão no Ténis e no Rugby. *Revista Portuguesa de Ciências do Desporto*, 6(3), 305-317.
- Pestana, M. H., & Gageiro, J. N. (2003). *Análise de Dados para Ciências Sociais. A complementaridade do SPSS* (3<sup>a</sup> ed.). Lisboa: Edições Sílabo.
- Phelps, R. P. (1999). Why testing experts hate testing? Testing student learning, evaluating teaching effectiveness. *Fordham Report*, 3(1). Acedido em 27 de Julho, 2011 em [www.eric.ed.gov](http://www.eric.ed.gov).
- Popham, W. J. (2003). *Test Better, Teach Better: The Instructional Role of Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Porcher, L. (2001). Note sur l'évaluation. In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 113-114). Porto Alegre: Artmed.
- Rezende, A., & Valdés, H. (2004). Métodos de estudo das habilidades táticas (3): inventários das habilidades esportivas. *Revista Digital – Buenos Aires*, 10(70). Acedido em 27 de Julho, 2011 em <http://www.efdeportes.com/>

- Rink, J., French, K., & Tjeerdsma, B. (1996). Foundations for the learning and instruction of sport and games. *Journal of Teaching in Physical Education*, 15(4), 399-417.
- Rink, J., Templeton, J., Hewitt, P., Dawkins, M., Mitchell, M., Barton, G., Taylor, M., & Hohn, R. (2002). High stakes assessment in South Carolina. *The Journal of Physical Education, Recreation & Dance*, 73(3) 21-24.
- Rodgers, E. G. (1937). An experimental investigation of the teaching of team games. *Teachers College Record*, 38(5), 435-437.
- Roldão, M. C. (1998). Que é ser professor hoje? – a profissionalidade docente revisitada. *Revista da ESES*, 1, Nova Série, 79-88.
- Roldão, M. C. (2000). O currículo escolar da uniformidade à contextualização – campos e níveis de decisão curricular. *Revista de Educação*, 9(1), 81-92.
- Roldão, M. C. (2001). A mudança anunciada da escola – ou a ruptura de um paradigma de escola? In I. Alarcão (Ed.), *Escola Reflexiva e Nova Racionalidade* (pp. 115-134). S. Paulo, Brasil: Artmed.
- Roldão, M. C. (2003). *Gestão do Currículo e Avaliação de Competências: As questões dos professores*. Lisboa: Editorial Presença.
- Roldão, M. C. (2005). Para um currículo do pensar e do agir: as competências enquanto referencial de ensino e aprendizagem. *Suplemento de En Direct de l' APPF*, Fev 2005, pp. 9-20.
- Roldão, M. C. (2006). Trabalho colaborativo – o que fazemos e o que não fazemos nas escolas? *Noesis*, 66, 22-23.
- Rovegno, I., Nevett, M., & Babiarz, M. (2001). Learning and teaching invasion-game tactics in 4<sup>th</sup> grade: Introduction and theoretical perspective. *Journal of Teaching in Physical Education*, 20(4), 341-351.
- Rovegno, I., Nevett, M., Brock, S., & Babiarz, M. (2001). Teaching and learning basic invasion-games tactics in 4<sup>th</sup> grade: A descriptive study from situated and constraints theoretical perspectives. *Journal of Teaching in Physical Education*, 20(4), 370-388.
- Scarloss, B. (2002). *Assessing complex academic performance at the group level*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, USA.
- Serpa, M. (2010). *Compreender a Avaliação: Fundamentos para Práticas Educativas*. Lisboa: Edições Colibri.
- Siedentop, D. (1991). *Developing Teaching Skills in Physical Education* (3th Ed.). Mountain View, California: Mayfield.

- Siedentop, D. (1994). *Sport education: Quality PE through positive sport experiences*. Champaign, IL: Human Kinetics.
- Siedentop, D., & Tannehill, D. (2000). *Developing Teaching Skills in Physical Education* (4<sup>a</sup> ed.). Mountain View, California: Mayfield.
- Silva, J. (2008). *Caracterização Técnico-Táctica de Jogos Reduzidos no Futebol*. Dissertação apresentada à Faculdade de Desporto da Universidade do Porto para a obtenção do grau de mestre, orientada por António Rebelo Natal, Porto.
- Sousa, A. B. (2005). *Investigação em Educação*. Lisboa: Livros Horizonte.
- Sternberg, R. J. (2008). Assessing What Matters. *Educational Leadership*, 65(4), 20-26.
- Stobart, G. (2010). *Assessment fit-for-the-future*. Paper presented at the IAEA Annual Conference, Bangkok, Thailand.
- Tabachnick, B., & Fidell, L. (1989). *Using multivariate statistics*. New York: Haper and Row Publishers.
- Tallir, I., Musch, E., Lannoo, K., & Van de Voorde, J. (2004). Validation of video-based instruments for the assessment of game performance in handball and soccer. In R. Light, K. Swabey, R. Brooker (Eds.). *Proceedings of the 2<sup>nd</sup> International Conference: Teaching Sport and Physical Education for Understanding* (pp. 108-113); Melbourne, Australia, 11 – 14 December, 2003. Melbourne, Australia: University of Melbourne.
- Tavares, F. (1998). O processamento da informação nos jogos desportivos. In A. Graça & J. Oliveira (Eds.), *O ensino dos jogos desportivos* (3<sup>a</sup> ed) (pp. 11-26). Porto: FCDEF.
- Téodorescu, L. (1977). *Théorie et methodologie des jeux sportifs*. Paris: Editions Français Réunis.
- Thomas, J., Nelson, J., & Silverman, S. (2007). *Métodos de Pesquisa em Atividade Física* (5<sup>a</sup> ed.) Porto Alegre: Artmed.
- Thorpe, R., & Bunker, D. (1986). Landmarks on our way to teaching for understanding. In R. Thorpe, D. Bunker, & L. Almond (Eds.), *Rethinking games teaching* (pp. 5-6). Loughborough: University of Technology.
- Tucker, P. D., & Stronge, J. H. (2005). *A avaliação dos professores e os resultados dos alunos* (Phala, Trad.). Porto. ASA Editores.
- Tuckman, B. (2000). *Manual de investigação em educação*. Lisboa: Fundação Calouste Gulbenkian.
- Turner, A., & Martinek, T. J. (1995). Teaching for Understanding: A model for improving decision making during game play. *Quest*, 47, 44-63.



- Van der Mars, H. (1989). Observer reliability: issues and procedures. In P. Darst, D. Zakrajsek & V. Mancini (Eds.), *Analyzing physical education and sport instruction* (pp. 53-79). Champaign, IL: Human Kinetics.
- Vial, M. (2001a). Instrumenter l'auto-evaluation – Contribution à la pensée complexe des faits d'éducation. In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 313-318). Porto Alegre: Artmed.
- Vial, M. (2001b). Un dispositif d'évaluation formatrice en expression écrite au college. In J. J. Bonniol & M. Vial (Eds.), *Modelos de Avaliação: Textos fundamentais* (pp. 204-205). Porto Alegre: Artmed.
- Wainer, H. (2007). The most dangerous equation. *American Scientist*, 95, 249-256.
- Webb, N. M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment*, 1(2), 131-152.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41-47.
- Wiggins, G. (1990). The Case for Authentic Assessment. *ERIC Digest*. Acedido em 15 de Julho de 2012 em <http://assessment.uconn.edu/>
- William, D. (2004). *Assessment and the regulation of learning*. Paper presented at Invited Symposium 'What does it mean for classroom assessment to be valid? Reliable?' at the annual meeting of the National Council on Measurement in Education, San Diego, CA, USA.
- Wood, T. M. (2003). Assessment in physical education: The future is now! In S. J. Silverman & C. D. Ennis (Eds.), *Student learning in physical education: Applying research to enhance instruction* (pp. 187-204). Champaign, IL: Human Kinetics.
- Zabalza, M. (1987). *Diseño y desarrollo curricular*. Madrid: Narcea.

## REFERÊNCIAS LEGISLATIVAS

- Decreto-Lei n. 6/2001 de 18 de Janeiro.
- Despacho Normativo n. 5/2007 de 10 de Janeiro.
- Despacho Normativo n. 1/2005 de 5 de Janeiro.



## Índice Remissivo Onomástico

### A

Abram, 117  
Afonso, 46  
Allal, 16, 33, 74  
Allen, 21, 123, 127, 132, 176  
Almond, 86, 87  
Ammons, 69  
Antunes, 90  
Araújo, 21, 100

### B

Babiarz, 105, 106, 115  
Barret, 38  
Barreto, 20, 22, 91, 101, 113, 137, 175  
Barroso, 19, 78  
Barton, 17, 48, 51  
Bastiaens, 18  
Batalau, 89  
Bayer, 88  
Behnke, 67, 68, 69, 118  
Bell, 86, 87  
Bengoechea, 94  
Bom, 15, 22, 31, 86, 112, 113, 115, 122, 137, 171  
Bonniol, 16, 29, 30, 32, 33, 34, 39, 48, 64, 139, 140  
Bouthier, 20, 21  
Bracken, 69, 118  
Braiuka, 96  
Bransgrove, 96  
Brighton, 36  
Brock, 106  
Brooker, 96  
Brooks, 69  
Bunker, 20, 92, 93, 175  
Butler, 87  
Button, 21

### C

Caldwell, 62  
Callahan, 36  
Campbell, 16, 35  
Campos, 18  
Cardinet, 54, 55, 63  
Carreiro da Costa, 22, 163, 167  
Carvalho, 15, 31  
Caty, 89

Caverni, 125, 134, 137, 138, 139, 140, 145, 153, 168  
Chow, 21, 99, 100, 175  
Clímaco, 18, 30, 32, 33, 39, 43, 44, 48, 52, 54, 59, 61, 77, 78, 123, 133, 169  
Cohen, 117, 118, 119  
Comédias, 15, 16  
Conceição, 18, 127  
Conselho Nacional das Associações de Professores e Profissionais de Educação Física, 16  
Coryn, 35, 39, 40  
Costa, 46  
Cothran, 83  
Coutinho, 136  
Crahay, 31, 46, 57, 59, 61, 62, 63, 68, 118  
Crespo, 15, 81  
Curtner-Smith, 97

### D

Darden, 150, 179  
Darling-Hammond, 19, 48, 52, 54, 55, 60, 79, 169  
Davids, 21, 100  
De Ketele, 18, 36, 37, 53, 64, 75, 130  
De Landsheere, 16, 38, 39, 41, 43, 55, 56, 57, 59, 61, 66, 74, 77, 128, 132  
Dejong, 81, 123  
Doering, 82, 83, 131  
Doolittle, 82, 85  
Dorfman, 37  
Dyson, 68, 95, 96, 97, 98, 113, 119, 121

### E

Eckes, 38  
Eisner, 72  
Elias, 104, 120, 174

### F

Fay, 82  
Fernandes, 16, 17, 18, 27, 28, 30, 31, 32, 33, 34, 35, 36, 40, 42, 44, 45, 46, 48, 49, 50, 52, 53, 54, 55, 56, 57, 58, 60, 61, 62, 63, 65, 66, 67, 70, 71, 74, 77, 78, 127, 130, 169  
Fidell, 136

Fleiss, 136  
Forster, 55  
Fragata, 39  
French, 113

## G

Gageiro, 159  
Gallagher, 43, 55  
Gardner, 16, 29, 30, 32, 36, 42, 44, 45, 62, 65, 74, 168  
Garganta, 15, 20, 88, 112, 115, 116, 117, 137  
Gaskin, 89, 94  
GEF-ESDLG, 24, 170, 171  
Genthon, 65  
Gerard, 18, 36, 37, 38, 53, 64, 66, 75, 130, 168  
Gipps, 36, 40  
Godbout, 20, 21, 85, 89, 108, 109  
Gonçalves, 89  
Goodson, 70  
Graça, 20, 21, 81, 84, 87, 89, 91, 94, 95, 96, 97, 98, 106, 113, 115, 175  
Greene, 55  
Gréhaigne, 19, 20, 21, 80, 81, 83, 84, 85, 88, 89, 90, 91, 94, 96, 102, 103, 106, 107, 108, 109, 112, 114, 115, 116, 123, 137, 150, 175  
Griffin, 19, 20, 21, 68, 81, 95, 175  
Gruber, 19  
Gulikers, 18, 37, 49, 62, 63, 64, 67, 70, 71, 72, 73, 77  
Guskey, 48, 54

## H

Hansen, 59, 68  
Hardman, 15  
Harlen, 16, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 53, 62, 65, 74, 77, 168, 169  
Harvey, 15, 20, 21, 83, 84, 90, 93, 94, 95, 96, 98, 102, 103, 104, 105, 107, 113, 114, 121, 137, 167, 175, 176  
Hastie, 68  
Hay, 19, 21, 79, 83, 84, 123  
Herrington, 17, 48, 58, 62, 64, 67, 73  
Hicks, 68  
Holt, 94, 95  
Hopper, 86, 87, 89, 91, 93, 94, 106  
Hristovski, 100  
Hubball, 85

## J

Jacinto, 15, 22, 86, 113, 122, 129  
Johnson, 19  
Júlio, 90, 100

## K

Kellaghan, 18, 38, 77  
King, 67, 68, 69, 118  
Kirk, 19, 21, 80, 93, 95, 96, 98, 114, 175  
Kirkendall, 19  
Kirschner, 18  
Kokinakis, 81  
Konzag, 88  
Kulinna, 83  
Kuntzleman, 81  
Kupermintz, 24, 151

## L

Lalanda, 18  
Lames, 15, 20, 112, 115, 116, 117, 137  
Lannoo, 21  
Liaw, 37  
Lotan, 117

## M

Madaus, 18, 38, 77  
Mahlo, 20, 90, 106, 175  
Mandigo, 87, 94  
Marion, 36, 37, 38  
Maroco, 157, 159  
Martin, 83, 85, 89, 94  
Martinek, 80  
Martins, 39  
McCaughtry, 97, 115  
McGarry, 15, 20, 112, 115, 116, 117, 137  
McMillan, 55, 56  
McPhail, 19, 21, 80, 94, 95, 98, 114, 175  
McTighe, 17, 30, 31, 32, 48, 58, 65, 66, 70, 75, 76  
Meisels, 37, 71  
Memmert, 21, 104, 105  
Mertler, 16, 35  
Mesquita, 20, 84, 87, 89, 90, 91, 94, 95, 96, 97, 98, 106, 112, 113, 115, 135  
Mira, 15  
Mitchell, 19, 20, 95  
Moon, 36, 38, 72, 74, 76, 131, 132  
Musch, 21

## **N**

Nadeau, 109  
National Association of State Boards of Education, 128  
Nazario, 16, 17, 18, 48, 55, 58, 64, 65, 71, 74, 77, 80, 81, 82, 84, 85, 112, 114, 117, 169  
Nelson, 24, 37  
Neves, 18  
Nevett, 105, 106, 115  
Newell, 99  
Noizet, 125, 134, 137, 138, 139, 140, 145, 153, 168

## **O**

O'Connor, 17, 30, 31, 32, 48, 58, 65, 66, 70, 75, 76  
Organisation for Economic Co-operation and Development, 34, 64, 65  
Oslin, 19, 20, 21, 80, 81, 88, 95, 102, 112, 114

## **P**

Pacheco, 19, 30, 32, 49, 54, 58, 59, 60, 61, 62, 65, 66, 78, 169  
Passos, 89, 100  
Pestana, 159  
Phelps, 56  
Piéron, 22, 163, 167  
Popham, 16, 17, 35, 37, 41, 42, 48, 50, 51, 54, 55, 57, 66, 69, 70, 71, 75, 76, 127, 128, 130, 131  
Porcher, 50

## **R**

Ratzlaff, 43, 55  
Renshaw, 21  
Rezende, 107  
Richard, 19, 109  
Rink, 42, 113  
Robison, 36  
Rodgers, 91  
Roldão, 17, 48, 49, 50, 56, 58, 59, 60, 61, 63, 65, 69, 78, 133, 169  
Rovegno, 97, 105, 106, 115

## **S**

Sampaio, 90

Scarloss, 117, 118, 119  
Schultz, 117  
Serpa, 38, 71, 73, 77, 122, 123  
Shuttleworth, 21  
Siedentop, 20, 24, 80, 82, 83, 84, 96, 114, 131, 175, 179  
Silva, 80, 101, 120  
Silverman, 24  
Sofo, 97  
Sousa, 138, 141, 145, 146, 150, 153  
Stephens, 59, 68  
Sternberg, 17, 44, 48, 57, 58, 60, 61, 62, 64, 66, 67, 68, 70, 75, 76  
Stobart, 17, 18, 36, 39, 40, 45, 55, 72, 176  
Strean, 94  
Stronge, 17, 48, 51, 52

## **T**

Tabachnick, 136  
Tallir, 21, 103, 107, 111, 121, 167  
Tannehill, 20, 80, 82, 83, 84, 114, 131  
Tavares, 106  
Téodorescu, 20, 88  
Thomas, 24, 125, 137, 139, 141, 144, 145, 150, 153, 160  
Thorpe, 20, 92, 93, 175  
Tjeerdsma, 113  
Tucker, 17, 48, 51, 52  
Tuckman, 24, 125, 141, 147, 153, 164  
Turner, 80

## **V**

Valdês, 107  
Van de Voorde, 21  
Van der Mars, 136  
Vial, 16, 29, 30, 32, 33, 39, 48, 64, 65

## **W**

Wainer, 52  
Webb, 119  
Wiggins, 18, 63, 72, 76, 169  
William, 16, 29, 34, 168  
Winters, 55  
Wood, 19, 20, 79, 82, 83, 85

## **Z**

Zabalza, 70

## **Apêndices**

## Apêndice 1

ESTAGIÁRIO ☐ PROFESSOR ☐ ANO CONCLUSÃO DO CURSO \_\_\_\_\_

É ESPECIALISTA EM JOGOS DESPORTIVOS COLECTIVOS? SIM ☐ NÃO ☐

EM QUE MODALIDADE É ESPECIALISTA \_\_\_\_\_

QUE FORMAÇÃO NESSA MODALIDADE:

ANOS DE PRÁTICA COMO JOGADOR \_\_\_\_\_

ANOS DE PRÁTICA COMO TREINADOR \_\_\_\_\_

CURSO DE TREINADOR \_\_\_\_\_

CURRÍCULO – ASPECTOS MAIS RELEVANTES \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

CONHECE O PROGRAMA DE ANDEBOL?

*NÃO, NUNCA LI* ☐

*LI OCASIONALMENTE* ☐

*CONSULTO FREQUENTEMENTE PARA AJUDAR-ME A TOMAR ALGUMAS DECISÕES* ☐

*UTILIZO PARA LECCIONAR E AVALIAR OS ALUNOS* ☐

-----

### JOGO 1

QUAL O NÍVEL DE JOGO? \_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

SE EXISTIR ALGUM JOGADOR DE NÍVEL DIFERENTE, QUAL É? \_\_\_\_\_

\_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## JOGO 2

QUAL O NÍVEL DE JOGO? \_\_\_\_\_

PORQUÊ? \_\_\_\_\_

---

---

---

SE EXISTIR ALGUM JOGADOR DE NÍVEL DIFERENTE, QUAL É? \_\_\_\_\_

---

PORQUÊ? \_\_\_\_\_

---

---

---

## JOGO 3

QUAL O NÍVEL DE JOGO? \_\_\_\_\_

PORQUÊ? \_\_\_\_\_

---

---

---

SE EXISTIR ALGUM JOGADOR DE NÍVEL DIFERENTE, QUAL É? \_\_\_\_\_

---

PORQUÊ? \_\_\_\_\_

---

---

---

#### **JOGO 4**

QUAL O NÍVEL DE JOGO? \_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

SE EXISTIR ALGUM JOGADOR DE NÍVEL DIFERENTE, QUAL É? \_\_\_\_\_

\_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

#### **JOGO 5**

QUAL O NÍVEL DE JOGO? \_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

SE EXISTIR ALGUM JOGADOR DE NÍVEL DIFERENTE, QUAL É? \_\_\_\_\_

\_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

## **JOGO 6**

QUAL O NÍVEL DE JOGO? \_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

SE EXISTIR ALGUM JOGADOR DE NÍVEL DIFERENTE, QUAL É? \_\_\_\_\_

\_\_\_\_\_

PORQUÊ? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Observações / comentários



## Apêndice 2

ESTAGIÁRIO ☐ PROFESSOR ☐ ANO CONCLUSÃO DO CURSO \_\_\_\_\_

É ESPECIALISTA EM JOGOS DESPORTIVOS COLECTIVOS? SIM ☐ NÃO ☐

EM QUE MODALIDADE É ESPECIALISTA? \_\_\_\_\_

QUE FORMAÇÃO NESSA MODALIDADE:

ANOS DE PRÁTICA COMO JOGADOR \_\_\_\_\_

ANOS DE PRÁTICA COMO ÁRBITRO \_\_\_\_\_

ANOS DE PRÁTICA COMO TREINADOR \_\_\_\_\_

CURSO DE TREINADOR \_\_\_\_\_

OUTRA ACTIVIDADE ESPECIALIZADA \_\_\_\_\_

OUTRA FORMAÇÃO ESPECIALIZADA \_\_\_\_\_

CURRÍCULUM – ASPECTOS MAIS RELEVANTES \_\_\_\_\_

---

---

---

---

---

### Apêndice 3

<b>JOGO 1</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>5</b> – Rapaz	Nº <b>2</b> – Rapaz
Nº <b>2</b> – Rapaz	Nº <b>3</b> – Rapaz
Nº <b>4</b> – Rapariga	Nº <b>1</b> – Rapariga
Nº <b>3</b> – Rapariga	Nº <b>4</b> – Rapariga

<b>JOGO 2</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>2</b> – Rapaz	Nº <b>1</b> – Rapaz
Nº <b>3</b> – Rapaz	Nº <b>8</b> – Rapaz
Nº <b>5</b> – Rapaz	Nº <b>5</b> – Rapariga
Nº <b>8</b> – Rapariga	Nº <b>3</b> – Rapariga

<b>JOGO 3</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>8</b> – Rapaz	Nº <b>4</b> – Rapaz
Nº <b>3</b> – Rapaz	Nº <b>5</b> – Rapaz
Nº <b>5</b> – Rapariga	Nº <b>3</b> – Rapariga
Nº <b>2</b> – Rapariga	Nº <b>2</b> – Rapariga

<b>JOGO 4</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>5</b> – Rapaz	Nº <b>3</b> – Rapaz
Nº <b>1</b> – Rapariga	Nº <b>8</b> – Rapariga
Nº <b>2</b> – Rapariga	Nº <b>5</b> – Rapariga
Nº <b>10</b> – Rapariga	Nº <b>4</b> – Rapariga

<b>JOGO 5</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>5</b> – Rapaz	Nº <b>4</b> – Rapaz
Nº <b>3</b> – Rapariga	Nº <b>5</b> – Rapaz
Nº <b>1</b> – Rapariga	Nº <b>2</b> – Rapariga
Nº <b>8</b> – Rapariga	Nº <b>7</b> – Rapariga

<b>JOGO 6</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>10</b> – Rapaz	Nº <b>3</b> – Rapaz
Nº <b>5</b> – Rapaz	Nº <b>7</b> – Rapaz
Nº <b>3</b> – Rapariga	Nº <b>1</b> – Rapariga
Nº <b>7</b> – Rapariga	Nº <b>5</b> – Rapariga

## Apêndice 4

<b>JOGO 1</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>3</b> – Rapariga	Nº <b>4</b> – Rapariga

<b>JOGO 2</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>8</b> – Rapariga	Nº <b>8</b> – Rapaz

<b>JOGO 3</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>2</b> – Rapariga	Nº <b>5</b> – Rapaz

<b>JOGO 4</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>1</b> – Rapariga	Nº <b>8</b> – Rapariga

<b>JOGO 5</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>1</b> – Rapariga	Nº <b>5</b> – Rapaz

<b>JOGO 6</b>	
<b>VERDES</b>	<b>LARANJAS</b>
Nº <b>7</b> – Rapariga	Nº <b>3</b> – Rapaz

## **Anexos**

## Protocolo de Avaliação dos JDC - ESDLG

### ANDEBOL

			Pág. 3
SITUAÇÃO de AVALIAÇÃO			
INTRODUÇÃO	ELEMENTAR	AVANÇADO	
5x5 (4+1x4+1) (campo reduzido e bola “afável”)	5x5 (4+1x4+1) ou 7x7 (campo reduzido)	7x7	
INDICADORES de OBSERVAÇÃO			
<ul style="list-style-type: none"> <li>* Há progressão no campo (a bola vai de “costa a costa”);</li> <li>* Não há aglomeração em torno da bola (“molhada”);</li> <li>* Há procura de espaço vazio à distância de passe (Desmarcação intencional);</li> <li>* Oportunidade de finalização para as duas equipas;</li> <li>* Há golos na sequência de remates em saltos não protagonizados (vários alunos);</li> </ul>	<ul style="list-style-type: none"> <li>* Jogo com largura e profundidade;</li> <li>* Há contra-ataques;</li> <li>* Muitas situações de 1x1 (fintas e remates apoiados e em suspensão);</li> <li>* Há capacidade ofensiva no 1x1;</li> <li>* Observa-se o trapézio ofensivo com circulação de bola (ocupação racional observada);</li> <li>* Há contacto físico (dentro das regras).</li> </ul>	<ul style="list-style-type: none"> <li>* Verifica-se o trapézio ofensivo;</li> <li>* Cria situações de superioridade numérica, tanto com posse de bola como sem a mesma, transição da defesa para o ataque;</li> <li>* Observam-se penetrações sucessivas e entradas;</li> <li>* Verifica-se defesa à zona 5:1 aberta;</li> <li>* Utiliza a técnica de remate mais adequada.</li> </ul>	

CRITÉRIOS de AVALIAÇÃO		
<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (início, recomeço, golos, foras, violações da área e dribles e passos, conduta para com adversário e penalizações), o objetivo do jogo e as principais ações técnico-táticas;</li> <li>* Boa pega de bola, passa com armação do braço;</li> <li>* “Quandos” do passe e do drible;</li> <li>* Remate em salto;</li> <li>* Desmarcação (“quebra do alinhamento”);</li> <li>* Reconhece-se como defesa, acompanha o adversário tentando impedir as ações do atacante;</li> <li>* <b>Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admitindo as falhas e opções dos colegas.</b></li> </ul>	<ul style="list-style-type: none"> <li>* Conhece as Regras Principais, o objetivo do jogo e as principais ações técnico-táticas, adequando a sua atuação a esse conhecimento, no papel de árbitro e jogador;</li> <li>* Desmarca-se rapidamente após recuperação da bola, ocupando o espaço de jogo de forma equilibrada;</li> <li>* Joga em profundidade e largura, garantindo a compensação ofensiva, trapézio ofensivo;</li> <li>* Ultrapassa adversário directo por exploração horizontal e vertical;</li> <li>* Colabora na circulação da bola, se não consegue vantagem numérica;</li> <li>* Como defesa, realiza MARCAÇÃO: individual, de vigilância e controlo.</li> <li>* <b>Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admitindo as falhas e opções dos colegas.</b></li> </ul>	<ul style="list-style-type: none"> <li>* Ocupa equilibradamente o espaço de jogo;</li> <li>* Conhece as movimentações que permitem a aquisição de superioridade numérica;</li> <li>* Bloqueio por parte do pivot;</li> <li>* Marcação de controlo, individual e de vigilância.</li> <li>* <b>Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admite as falhas dos colegas e dá sugestões que favoreçam a sua melhoria.</b></li> </ul>

## ANDEBOL

SITUAÇÃO de AVALIAÇÃO		
N+	I+	E+
5x5 (4+1x4+1) (campo reduzido e bola “afável”)	5x5 (4+1x4+1) ou 7x7 (campo reduzido)	7x7
INDICADORES de OBSERVAÇÃO		
<ul style="list-style-type: none"> <li>* Há progressão no campo (a bola vai de “costa a costa”);</li> <li>* Há procura de espaço vazio à distância de passe (Desmarcações).</li> </ul>	<ul style="list-style-type: none"> <li>* Há muita posse de bola;</li> <li>* Muitos remates;</li> <li>* Mudança de atitude após perda da posse de bola ou recuperação da mesma;</li> <li>* Defesas entre atacante e baliza;</li> <li>* Capacidade ofensiva no 1x1;</li> <li>* Observa-se o trapézio ofensivo.</li> </ul>	<ul style="list-style-type: none"> <li>* Observa-se nitidamente as fases de jogo particularmente contra-ataque, organização do ataque recuperação defensiva e organização em situação defensiva;</li> <li>* Observa-se em situação 5:1 com posições abertas;</li> <li>* Ocupação do posto específico;</li> <li>* Observa-se os meios táticos de grupo ofensivos (penetrações sucessivas, cruzamentos e entradas);</li> <li>* Observam-se os meios táticos defensivos de grupo: troca de posições e adversários; ajuda.</li> </ul>



CRITÉRIOS de AVALIAÇÃO		
<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (início, recomeço, golos, foras, violações da área e dribles e passos, conduta para com adversário e penalizações), o objetivo do jogo e as principais ações técnico-táticas;</li> <li>* Pega e Armação do braço;</li> <li>* “Quandos” do passe e do drible;</li> <li>* Desmarcação (“quebra do alinhamento”);</li> <li>* Reconhece-se como defesa e acompanha adversário.</li> </ul>	<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (início, recomeço, golos, foras, violações da área e dribles e passos, conduta para com adversário e penalizações), o objetivo do jogo e as principais ações técnico-táticas;</li> <li>* Desmarca-se rapidamente após recuperação da bola;</li> <li>* Ocupa equilibradamente o espaço de jogo;</li> <li>* Ultrapassa adversário directo por exploração horizontal e vertical;</li> <li>* Colabora na circulação da bola e faz compensação ofensiva;</li> <li>* Como defesa, realiza MARCAÇÃO individual;</li> <li>* 1 x 1: finta e remate á frente do defensor.</li> </ul>	

**NOTA:** Em caso de duvida sobre determinado aluno, pelo facto de não ser possível avaliá-lo uma vez que o mesmo não teve oportunidade de realizar (em jogo) a acção exigida, deve-se realizar um exercicio critério que permita ao avaliador observar se o aluno efectua ou não determinada acção.

## FUTEBOL

SITUAÇÃO de AVALIAÇÃO		
INTRODUÇÃO	ELEMENTAR	AVANÇADO
4x4 (Gr+3x3+Gr) ou 3+1x3+1 (espaço amplo)	5x5 (Gr+4x4+Gr) ou 4+1x4+1 (campo normal)	7x7 (Gr+6x6+Gr) ou 11x11 (Gr+10x10+Gr)
INDICADORES de OBSERVAÇÃO		
<ul style="list-style-type: none"> <li>* Recepção orientada;</li> <li>* A bola está muito em campo;</li> <li>* Não há aglomeração em torno da bola (“molhada”);</li> <li>* Há procura de espaço vazio à distância de passe (Desmarcação intencional);</li> <li>* Há progressão no campo (a bola vai de “costa a costa”);</li> <li>* Observam-se remates.</li> <li>* Oportunidade de finalização para as duas equipas.</li> </ul>	<ul style="list-style-type: none"> <li>* Capacidade de manter a bola quando está marcado;</li> <li>* Há muita posse de bola;</li> <li>* Disposição jogadores no campo;</li> <li>* Exploração de superioridade numérica ofensiva;</li> <li>* Jogo com largura e profundidade;</li> <li>* Qualidade nas execuções;</li> <li>* Muitos remates.</li> <li>* Defesas entre o atacante e a baliza;</li> </ul>	<ul style="list-style-type: none"> <li>* Constrói acções claras ofensivas de superioridade numérica;</li> <li>* Desmarcações de apoio e de ruptura;</li> <li>* Realiza marcação à zona;</li> <li>* Centro de jogo definido;</li> <li>* Conhece os princípios gerais do jogo: procura superioridade numérica, evita a igualdade numérica e recusa a inferioridade;</li> <li>* Utilização adequada dos três corredores de jogo, desorganizar a estrutura defensiva adversária;</li> <li>* Princípios de jogo evidentes.</li> </ul>

CRITÉRIOS de AVALIAÇÃO		
<p>* Conhece as Regras Principais (início, recomeço, golos, foras, cantos, faltas e penalty) e o objectivo do jogo;</p> <p>* Há procura de espaço vazio à distância de passe (desmarcação) aclarando o jogo;</p> <p>* Progressão: para rematar (se tem espaço) ou passar (se colega em melhor posição);</p> <p>* Quando em situação, remata;</p> <p>* "Quandos" do passe e da condução;</p> <p>* Se não for pressionado:                      . Recepção enquadrada;                      . Passe a colega desmarcado;</p> <p>* Mudança de atitude rápida, ofensiva para defensiva e vice versa;</p> <p>* Reconhece-se como defesa e acompanha adversário.</p> <p><b>* Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admitindo as opções e falhas dos colegas.</b></p>	<p>* Conhece as Regras Principais (início, recomeço, golos, foras, cantos, faltas e penalty) e o objectivo do jogo;</p> <p>* Recepção com enquadramento ofensivo: passe (se colega desmarcado perto da baliza), remate (se próximo da baliza), progressão para rematar passar ou fintar;</p> <p>* Desmarcação em função da baliza (ruptura/apoio);</p> <p>* Marcação ao adversário para dificultar a acção ofensiva;</p> <p>* Muda de atitude, ofensiva e defensiva em função do momento de jogo, colocando-se entre a bola e a baliza e intercepta linhas de passe;</p> <p>* Capacidade de manter a bola quando marcado;</p> <p>* Conforme situação remata, passa ou conduz a bola com qualidade;</p> <p>* Aclaramento à desmarcação.</p> <p><b>* Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admitindo as opções e falhas dos colegas.</b></p>	<p>* Acções técnico-táticas individuais e colectivas garantindo segurança no ataque: combinações táticas (simples/diretas e indirectas), penetrações, fintas e desmarcações;</p> <p>. Acções técnico-táticas defensivas: dobras, defesa à zona, princípios defensivos (contenção, apoio defensivo, concentração), pressão ao portador da bola quando este se aproxima da baliza, marcação individual e à zona;</p> <p>. Muda de atitude, ofensiva e defensiva em função do momento de jogo, colocando-se rapidamente entre a bola e a baliza e intercepta linhas de passe.</p> <p><b>* Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admite as falhas dos colegas e dá sugestões que favoreçam a sua melhoria.</b></p>

## FUTEBOL

SITUAÇÃO de AVALIAÇÃO		
N+	I+	E+
4x4 (Gr+3x3+Gr) ou 3+1x3+1 (espaço amplo)	5x5 (Gr+4x4+Gr) ou 4+1x4+1 (campo normal)	5x5
INDICADORES de OBSERVAÇÃO		
<ul style="list-style-type: none"> <li>* Poucos passes sucessivos;</li> <li>* Não há aglomeração em torno da bola (“molhada”);</li> <li>* Jogadores tentam controlar a bola.</li> </ul>	<ul style="list-style-type: none"> <li>* Capacidade de manter a bola quando está marcado;</li> <li>* Há muita posse de bola;</li> <li>* Qualidade nas execuções;</li> <li>* Muda de atitude, ofensiva e defensiva em função do momento de jogo;</li> <li>* Conforme a situação remata, passa ou conduz a bola com qualidade controlada;</li> <li>* Muitos remates;</li> <li>* Defesas entre atacante e baliza.</li> </ul>	<ul style="list-style-type: none"> <li>* Garantia de ações ofensivas com profundidade e largura;</li> <li>* Vêm-se movimentos 1x1 de penetração;</li> <li>* Criação de situações de superioridade numérica ofensiva;</li> <li>* Observa-se contenção, apoio defensivo e marcação do jogador com e sem bola;</li> <li>* Utilização adequada dos três corredores de jogo.</li> </ul>

CRITÉRIOS de AVALIAÇÃO		
<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (início, recomeço, golos, foras, cantos, faltas e penalty) e o objectivo do Jogo;</li> <li>* Se não for pressionado, tenta:                         <ul style="list-style-type: none"> <li>. Recepção enquadrada;</li> <li>. Passe a colega desmarcado;</li> </ul> </li> <li>* Há desmarcações;</li> <li>* Quando em situação, remata;</li> <li>* Reconhece-se como defesa.</li> </ul>	<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (início, recomeço, golos, foras, cantos, faltas e penalty), o objectivo do Jogo e as principais acções técnico-táticas;</li> <li>* Recepção com enquadramento ofensivo;</li> <li>* Conforme a situação remata, passa ou conduz a bola com qualidade controlada;</li> <li>* Desmarcações (ofensivas) em função do alvo (baliza);</li> <li>* Ataque em largura e profundidade;</li> <li>* Capacidade de manter a bola quando marcado;</li> <li>* Marca o adversário.</li> </ul>	

**NOTA:** Em caso de duvida sobre determinado aluno, pelo facto de não ser possível avaliá-lo uma vez que o mesmo não teve oportunidade de realizar (em jogo) a acção exigida, deve-se realizar um exercicio critério que permita ao avaliador observar se o aluno efectua ou não determinada acção.

## BASQUETEBOL

		Pág. 1
SITUAÇÃO de AVALIAÇÃO		
INTRODUÇÃO	ELEMENTAR	AVANÇADO
<ul style="list-style-type: none"> <li>* 3x3 (campo reduzido com duas tabelas)</li> <li>* 4x4 (campo reduzido)</li> </ul>	5x5 (também se pode ver no 4x4)	5X5
INDICADORES de OBSERVAÇÃO		
<ul style="list-style-type: none"> <li>* Há progressão no campo (a bola vai de “costa a costa”);</li> <li>* Não há aglomeração em torno da bola (“molhada”);</li> <li>* Há procura de espaço vazio à distância de passe (Desmarcação intencional) para receber a bola;</li> <li>* Oportunidade de finalização para as duas equipas;</li> <li>* Observam-se muitos lançamentos, na passada, não protagonizados (vários alunos).</li> </ul>	<ul style="list-style-type: none"> <li>* Na transição defesa/ataque condução de bola preferencialmente no corredor central;</li> <li>* Jogadores em posição base defensiva, entre atacante e cesto;</li> <li>* Atitude defensiva (dificulta o passe e a progressão do atacante);</li> <li>* Observa-se os “5 abertos”;</li> <li>* Movimentos ofensivos claros na direcção do cesto;</li> <li>* Muita luta nos ressaltos (ofensivo e defensivo);</li> <li>* Muitas situações de 1x1;</li> <li>* Capacidade ofensiva no 1x1;</li> <li>* Muitos lançamentos na passada.</li> <li>* Capacidade de mudança de mão;</li> </ul>	<ul style="list-style-type: none"> <li>* Exploração de situações de superioridade numérica;</li> <li>* Desmarcação intencional de modo a permitir o ataque em cinco abertos;</li> <li>* Corte para o cesto após passe;</li> <li>* Atitude de tripla ameaça;</li> <li>* Dá continuidade ao ressalto ofensivo;</li> <li>* Ressalto defensivo;</li> <li>* Preocupações defensivas.</li> </ul>

CRITÉRIOS de AVALIAÇÃO		
<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (formas de jogar a bola, início e recomeço, foras, dribles e passos, bola presa, faltas pessoais, três segundos), o objetivo do jogo e as principais ações técnico-táticas;</li> <li>* Enquadramento ofensivo;</li> <li>* "Quandos" do Passe e do Drible;</li> <li>* Conforme situação lança parado ou na passada;</li> <li>* Desmarcação ofensiva (criando linhas de passe e ocupando racionalmente o espaço de jogo);</li> <li>* Quando a equipa perde a bola assume atitude defensiva, marcação individual;</li> <li>* Participa no ressalto sempre que há lançamento;</li> <li>* <b>Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admitindo as falhas e opções dos colegas.</b></li> </ul>	<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (formas de jogar a bola, início e recomeço, foras, dribles e passos, bola presa, faltas pessoais, três segundos), o objetivo do jogo e as principais ações técnico-táticas;</li> <li>* Trabalho de receção;</li> <li>* Reposição;</li> <li>* Aclarar ao drible, Passe e Corte;</li> <li>* Libertação do defensor directo;</li> <li>* Marcação;</li> <li>* Participa no ressalto defensivo e ofensivo;</li> <li>* Colocação do defesa entre a bola e o cesto, a fim, de dificultar o passe, drible ou lançamento do adversário;</li> <li>* <b>Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admite as falhas dos colegas e dá sugestões que favoreçam a sua melhoria.</b></li> </ul>	<ul style="list-style-type: none"> <li>* Aproveitamento de situações de vantagem;</li> <li>* Imprevisibilidade no ataque;</li> <li>* Continuidade das acções ofensivas, se não existe aproveitamento das situações de superioridade numérica ou se não recebe a bola após corte na direcção do cesto após passe.</li> <li>* <b>Coopera com os companheiros, promovendo o êxito pessoal e do grupo, admite as falhas dos colegas e dá sugestões que favoreçam a sua melhoria.</b></li> </ul>

## BASQUETEBOL

SITUAÇÃO de AVALIAÇÃO		
N+	I+	E+
* 3x3 (campo reduzido com duas tabelas) * 4x4 (campo reduzido)	5x5 (também se pode ver no 4x4)	5x5
INDICADORES de OBSERVAÇÃO		
* Há progressão no campo (a bola vai de “costa a costa”);  * Há deslocamentos no campo (em progressão);  * Há muita procura do espaço vazio e à distância de passe;  * Oportunidade de finalização para as duas equipas.	* Muitas situações de 1x1;  * Capacidade ofensiva no 1x1; * Muitos lançamentos na passada; * Atitude defensiva (dificulta o passe e a progressão do atacante); * Movimentos ofensivos claros na direcção do cesto;  * Muita luta nos ressaltos;  * Jogadores em posição base defensiva, entre atacante e cesto.	* Drible aproximando a bola do cesto, procurando linha de passe;  * Observam-se ataques em 5 abertos; * Vêm-se alunos numa atitude de tripla ameaça;  * Observam-se cortes, após passe penetrante;  * Observa-se sobremarcação e corte na sobremarcação.



CRITÉRIOS de AVALIAÇÃO		
<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (formas de jogar a bola, início e recomeço, foras, dribles e passos, bola presa, faltas pessoais, três segundos) e o Objectivo do Jogo;</li> <li>* Enquadramento ofensivo;</li> <li>* “Quandos” do passe e do drible;</li> <li>* Desmarcação ofensiva;</li> <li>* Reconhece-se como defesa.</li> </ul>	<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (formas de jogar a bola, início e recomeço, foras, dribles e passos, bola presa, faltas pessoais, três segundos), as principais acções técnico-táticas e o Objectivo do Jogo;</li> <li>* Trabalho de receção;</li> <li>* Libertação do defensor direto;</li> <li>* Marcação;</li> <li>* Ressalto defensivo e ofensivo;</li> <li>* Aclarar ao drible;</li> <li>* Passe e corte;</li> <li>* Marcação;</li> <li>* Colocação do defesa entre a bola e o cesto.</li> </ul>	

**NOTA:** Em caso de duvida sobre determinado aluno, pelo facto de não ser possível avaliá-lo uma vez que o mesmo não teve oportunidade de realizar (em jogo) a acção exigida, deve-se realizar um exercicio critério que permita ao avaliador observar se o aluno efectua ou não determinada acção.

## VOLEIBOL

Pág. 7

SITUAÇÃO de AVALIAÇÃO		
INTRODUÇÃO	ELEMENTAR	AVANÇADO
2+2 com serviço - Rede a 2m Serviço 3 a 4.5m	4x4 (campo reduzido 12m x 6m) Rede a 2.10/2.15	6x6 (jogo formal)
INDICADORES de OBSERVAÇÃO		
<p>* Bola anda muito tempo no ar (passa várias vezes a rede sem ser ping-pong);</p> <p>* A maioria dos passes é em toque dedos;</p> <p>* Mínimo 2 toques, máximo 3.</p>	<p>* Observa-se a dinâmica dos três toques:</p> <ul style="list-style-type: none"> <li>ü Há alguém que recebe;</li> <li>ü Há alguém que passa;</li> <li>ü Finalização intencional (dificultando a ação do adversário).</li> </ul> <p>* 2º toque maior parte das vezes na posição 3;</p> <p>* Observam-se finalizações intencionais;</p> <p>* Qualidade nas execuções (passe, manchete, serviço, remate).</p>	<p>* Existe dinâmica dos três toques:</p> <ul style="list-style-type: none"> <li>ü Há recepção direccionada ao passador;</li> <li>ü Há um passador</li> <li>ü Há finalização intencional;</li> </ul> <p>* Oposição ao ataque;</p> <p>* Comunica com os companheiros e posiciona-se intencionalmente para receber o serviço (manchete ou por cima).</p>

CRITÉRIOS de AVALIAÇÃO		
<p>* Conhece as Regras Principais (2 toques, toques consecutivos, transporte, rotação) e o objectivo do jogo;</p> <p>* Colocação no ponto de queda da bola (com deslocamento), com recepção a 2 mãos por cima ou em manchete;</p> <p>* Serviço por baixo direccionado;</p> <p>* Toque (dedos ou antebraços) para dar continuidade ao jogo.</p> <p><b>* Cooperar com os companheiros, promovendo o êxito pessoal e do grupo, admitindo as opções e falhas dos colegas.</b></p>	<p>* Conhece as Regras Principais (2 toques, toques consecutivos, transporte, rotação) e o objectivo do jogo;</p> <p>* Recepção dirigida para um companheiro;</p> <p>* Passe intencional;</p> <p>* Finalização em remate (em apoio) ou passe colocado no espaço vazio (que dificulte recepção);</p> <p>* Recepção (passe ou manchete) para dar continuidade ao jogo;</p> <p>* Passe para finalização (2º toque);</p> <p>* Serviço por baixo ou por cima (ténis) direccionado (para zona de difícil recepção).</p> <p><b>* Cooperar com os companheiros, promovendo o êxito pessoal e do grupo, admitindo as opções e falhas dos colegas.</b></p>	<p>* Recepção com qualidade;</p> <p>* Passe com qualidade;</p> <p>* Remate, amorti ou passe colocado de acordo com o posicionamento do adversário;</p> <p>* Realiza bloco e protecção ao mesmo;</p> <p>* Ao remate da sua equipa, avança no terreno, assumindo atitude de protecção ao ataque;</p> <p>* Serviço por baixo ou por cima direccionado, colocando a bola no meio-campo oposto em condições de difícil recepção.</p> <p><b>* Cooperar com os companheiros, escolhendo as acções favoráveis à vantagem da sua equipa, aceitando as indicações que lhe dirigem, bem como as opções e falhas dos seus colegas. Analisa a sua prestação e a dos companheiros, dando indicações e sugestões que favoreçam a sua melhoria.</b></p>

## VOLEIBOL

Pág. 8

SITUAÇÃO de AVALIAÇÃO		
N+	I+	E+
2+2 com serviço em passe Rede a 2m	4x4 Rede a 2.10/2.15 Campo (12mx6m)	4x4 ou 6x6
INDICADORES de OBSERVAÇÃO		
<p>* Bola anda muito tempo no ar (passa várias vezes a rede sem ser ping-pong);</p> <p>* A maioria dos passes é em toque dedos;</p> <p>* Grande maioria dos passes altos de frente.</p>	<p>* Existe dinâmica dos três toques: ü Há alguém que recebe ü Há alguém que passa</p> <p>* 2º toque maior parte das vezes na posição 3;</p> <p>* Qualidade nas execuções (passe, manchete, serviço).</p> <p>Qualidade nas execuções (passe, manchete, serviço)</p>	<p>* Observa-se uma organização coletiva: . para recepção ao serviço, em w (4:0:2); . para defesa ao ataque adversário (em 3:1:2); . para proteção ao ataque da sua equipa em 2 linhas (6 avançados);</p> <p>* Observam-se muitos remates;</p> <p>* Bloco a 2.</p>

CRITÉRIOS de AVALIAÇÃO		
<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (2 toques, toques consecutivos, transporte, rotação);</li> <li>* Colocação no ponto de queda da bola (com deslocamento);</li> <li>* Toque (dedos ou antebraços) para dar continuidade ao jogo.</li> </ul>	<ul style="list-style-type: none"> <li>* Conhece as Regras Principais (2 toques, toques consecutivos, transporte, rotação);</li> <li>* Passe colocado;</li> <li>* Serviço por baixo ou por cima (ténis) direccionado;</li> <li>* Recepção com qualidade (passe ou manchete) para dar continuidade ao jogo;</li> <li>* Passe para finalização (2º toque);</li> <li>* Serviço por baixo ou por cima (de ténis) para zona de difícil recepção.</li> </ul>	

**NOTA:** Em caso de duvida sobre determinado aluno, pelo facto de não ser possível avaliá-lo uma vez que o mesmo não teve oportunidade de realizar (em jogo) a acção exigida, deve-se realizar um exercicio critério que permita ao avaliador observar se o aluno efectua ou não determinada acção.